Multi-Modal Activity Recognition Systems With Minimal Training Data and Unobtrusive Environmental Instrumentations

Thesis approved by the Department of Computer Science, Technical University of Kaiserslautern for the award of the doctoral degree:

Doctor of Engineering (Dr.-Ing.)

Presented by:
Gerald Bauer

Affiliation:
Embedded Intelligence Group
German Research Center for Artificial Intelligence (DFKI)

Dean: Prof. Dr. Arnd Poetzsch-Heffter
Chair of committee: Prof. Dr. Katharina Zweig
Thesis examiner: Prof. Dr. Paul Lukowicz
Thesis co-examiner: Prof. Dr. Bernt Schiele

Date of Defense: 19 December 2013

D386

Multi-Modal Activity Recognition Systems With Minimal Training Data and Unobtrusive Environmental Instrumentations

Gerald Bauer

PhD thesis

Abstract

The recognition of day-to-day activities is still a very challenging and important research topic. During recent years, a lot of research has gone into designing and realizing smart environments in different application areas such as health care, maintenance, sports or smart homes. As a result, a large amount of sensor modalities were developed, different types of activity and context recognition services were implemented and the resulting systems were benchmarked using state-of-the-art evaluation techniques. However, so far hardly any of these approaches have found their way into the market and consequently into the homes of real end-users on a large scale. The reason for this is, that almost all systems have one or more of the following characteristics in common: expensive high-end or prototype sensors are used which are not affordable or reliable enough for mainstream applications; many systems are deployed in highly instrumented environments or so-called "living labs", which are far from real-life scenarios and are often evaluated only in research labs; almost all systems are based on complex system configurations and/or extensive training data sets, which means that a large amount of data must be collected in order to install the system. Furthermore, many systems rely on a user and/or environment dependent training, which makes it even more difficult to install them on a large scale. Besides, a standardized integration procedure for the deployment of services in existing environments and smart homes has still not been defined. As a matter of fact, service providers use their own closed systems, which are not compatible with other systems, services or sensors. It is clear, that these points make it nearly impossible to deploy activity recognition systems in a real daily-life environment, to make them affordable for real users and to deploy them in hundreds or thousands of different homes.

This thesis works towards the solution of the above mentioned problems. Activity and context recognition systems designed for large-scale deployment and real-life scenarios are introduced. Systems are based on low-cost, reliable sensors and can be set up, configured and trained with little effort, even by technical laymen. It is because of these characteristics that we call our approach "minimally invasive". As a consequence, large amounts of training data, that are usually required by many state-of-the-art approaches, are not necessary. Furthermore, all systems were integrated unobtrusively in real-world/similar to real-world environments and were evaluated under real-life, as well as similar to real-life conditions. The thesis addresses the following topics: First, a sub-room level indoor positioning system is introduced. The system is based on low-cost ceiling cameras and a simple computer vision tracking approach. The problem of user identification is solved by correlating modes of locomotion patterns derived from the trajectory of unidentified objects and on-body motion sensors. Afterwards, the issue of recognizing *how* and *what* mainstream household devices have been used *for* is considered. Based on a low-cost microphone, the water consumption of water-taps can be approximated by analyzing plumbing noise. Besides that, operating modes of mainstream electronic devices were recognized by using rule-based classifiers, electric current features and power measurement sensors. As a next step, the difficulty of spotting subtle, barely distinguishable hand activities and the resulting object interactions, within a data set containing a large amount of background data, is addressed. The problem is solved by introducing an on-body core system which is configured by simple, one-time physical measurements and minimal data collections. The lack of large training sets is compensated by fusing the system with activity and context recognition systems, that are able to reduce the search space observed. Amongst other systems, previously introduced approaches and ideas are revisited in this section. An in-depth evaluation shows the impact of each fusion procedure on the performance and run-time of the system. The approaches introduced are able to provide significantly better results than a state-of-the-art inertial system using large amounts of training data. The idea of using unobtrusive sensors has also been applied to the field of behavior analysis. Integrated smartphone sensors are used to detect behavioral changes of individuals due to medium-term stress periods. Behavioral parameters related to location traces, social interactions and phone usage were analyzed to detect significant behavioral changes of individuals during stressless and stressful time periods. Finally, as a closing part of the thesis, a standardization approach related to the integration of ambient intelligence systems (as introduced in this thesis) in real-life and large-scale scenarios is shown.

Zusammenfassung

Die automatische Erkennung von Personenaktivitäten ist ein äußerst herausforderndes Forschungsthema, dem vor allem in praxisrelevanten Szenarien großes Interesse gewidmet wird. Viele Arbeiten beschäftigten sich in den letzten Jahren mit der Entwicklung von intelligenten Systemen für verschiedenste Anwendungsbereiche, wie zum Beispiel Instandhaltung, Gesundheitswesen oder intelligentes Wohnen. So wurden eine Vielzahl von Sensormodalitäten, Aktivitäts- und Kontexterkennungssysteme entwickelt und evaluiert. Jedoch konnten bis dato nur sehr wenige dieser Systeme auf den Markt gebracht und somit grossflächig in die Wohnungen von Endanwendern integriert werden. Der Grund hierfür liegt darin begründet, dass beinahe alle Systeme die folgenden Merkmale aufweisen: Zum einen wurden meist kostspielige High-End Sensoren oder Prototypen verwendet, die nicht ausreichend zuverlässig oder aber für den Durchschnittsverdiener nicht erschwinglich sind. Darüber hinaus wurden viele Systeme in sensorreichen Umgebungen, den sogenannten "Living Labs", oder in Forschungslaboren getestet, welche mit realen Umgebungen nicht vergleichbar sind. Zudem können viele Systeme nur anhand komplexer Konfigurationsprozesse oder auf Basis von großen Trainingsdatenmengen initialisiert werden. Des Weiteren müssen viele Systeme an ihre Umgebung oder an den Nutzer angepasst werden, was wiederum bei großflächigen Einsätzen nur schwer realisierbar ist. Letztendlich fehlt ein allgemein anerkanntes Standardisierungsverfahren für die einheitliche Integration in intelligente Umgebungen. Aufgrund dieser Fakten ist es schwierig, solche Systeme großflächig in reale Alltagsumgebungen zu integrieren und zu einem erschwinglichen Preis anzubieten.

Diese Arbeit setzt sich mit Lösungsansätzen für die oben genannten Probleme auseinander. Es werden Aktivitäts- und Kontexterkennungssysteme für den großflächigen Einsatz in realen Anwendungsszenarien vorgestellt. Die entwickelten Konzepte basieren auf erschwinglichen, zuverlässigen Sensoren und können durch einfache Konfigurations- und Trainingsprozesse auch von weniger technikaffinen Personen genutzt werden. Aus diesen Gründen wird der hier vorgestellte Ansatz als "minimal invasiv" bezeichnet. Folglich werden große Mengen von Trainingsdaten, wie sie von vielen State-of-the-Art Ansätzen verwendet werden, nicht benötigt. Des Weiteren wurden die gezeigten Systeme auf eine "unaufdringliche" Art in reale / realitätsnahe Szenarien integriert und evaluiert. Im Detail befasst sich diese Arbeit mit den folgenden Themen: Als Erstes wird die Thematik der Positionsbestimmung von Personen innerhalb geschlossener Räume betrachtet. Anhand von Deckenkameras und einfachen Bildverarbeitungsalgorithmen können Personenbewegungen erkannt und lokalisiert werden. Die vorherrschende Problematik der Personenidentifikation wird durch die Korrelation von Bewegungsmustern auf Basis von Bewegungssensoren, die Personen zugeordnet sind, und detektierten Objekttrajektorien gelöst. Anschließend wird das Problem der automatischen Erkennung der Nutzung von Haushaltsgeräten behandelt. Zum einen wird der Wasserkonsum einzelner Wasserhähne anhand der Audioanalyse von Wassergeräuschen in Zufluss-Rohren approximiert. Zusätzlich wird ein System vorgestellt, welches basierend auf der Analyse von Stromkennzahlen nicht nur die Nutzung von elektronischen Geräten, sondern auch deren Betriebsmodi erkennen kann. Im nächsten Schritt werden die vorgestellten Konzepte genutzt, um subtile, schwer unterscheidbare Handaktivitäten und die daraus resultierenden Objektinteraktionen in einer großen Menge von Alltagsaktivitäten zu detektieren und zu klassifizieren. Das Problem wird auf Basis eines tragbaren Systems gelöst, welches durch einfache und einmalig auszuführende physikalische Messungen und minimale Datenaufnahmen konfiguriert wird. Das Fehlen von großen Trainingsdatenmengen wird durch die Fusion von mehreren autarken Aktivitätserkennungssystemen, welche den betrachteten Suchraum stark einschränken, kompensiert. Es wird gezeigt, dass das vorgestellte System signifikant bessere Resultate erzielt als ein State-of-the-Art Ansatz basierend auf großen Trainingsdatenmengen. Die Nutzung von unaufdringlichen Sensoren wird zudem auf den Bereich der Verhaltensanalyse übertragen. Smartphone-Sensoren werden genutzt, um Verhaltensänderungen von Individuen aufgrund von mittelfristigen Stressperioden zu erkennen. Dafür werden Verhaltensmuster in Bezug auf geographische Aufenthaltsorte, soziale Interaktionen und der Mobiltelefonnutzung berechnet und analysiert. Im abschließenden Teil der Arbeit wird ein Standardisierungsansatz vorgestellt, mit dem die Integration von "Ambient Intelligence" Systemen – wie sie in dieser Arbeit gezeigt wurden – in reale und großflächige Szenarien ermöglicht wird.

Acknowledgments

First of all, I would like to express my deep gratitude to my thesis advisor Prof. Dr. Paul Lukowicz for giving me the opportunity to work in his research groups at the University of Passau and the German Research Center of Artificial Intelligence (DFKI). I thank him for his valuable research input, guidance and excellent inspiration over the past years as well as for the possibility of working in outstanding research projects. I could always rely on his enormous experience and his advice.

I am very proud to have Prof. Dr. Bernt Schiele as my second advisor. He is a brilliant researcher and I would like to say special thanks to him and his research group at the Max Planck Institute in Saarbrücken for fruitful and valuable discussions and interesting research collaborations during the last few years.

My gratitude goes to Tobias Franke, Kamil Kloch and Gernot Bahle for endless helpful discussions, being great roommates and friends. I would like to thank Jens Weppner, Gerald Pirkl, Dr. Jingyuan Cheng, Agnes Grünerbl, Dr. George Kampis, Karl Stockinger, David Bannach, Dr. Georg Ogris, Dr. Kai Kunze and Jane Bensch for being great colleagues.

I would especially like to thank Bruno Jean-Bart, Markus Dubielzig, Oskar Jonsson, Margaret Ellis and Prof. Dr. Dusan Simsik for being great friends as well as excellent project partners and for sharing their expertise. Without these people the MonAMI project would not have been so exciting and successful.

A big personal thank you goes to Prof. Dr. Armando Roy, Dr. Roberto Casas, Ruben Blasco Marin, Luis Lain, Alejandro Ibarz García, Dr. Alvaro Marco, Alba Jimenez, their families and all members of Tecnodiscap for great research projects, a great working atmosphere at the University of Zaragoza and the amazing and unforgettable time in Zaragoza.

Warm thanks go to all my friends and especially to Sandra Reitbauer and Johannes Götz for their encouragement and just for being there.

Most importantly, I sincerely thank my parents, Maria and Gerhard Bauer, my sister Sabrina and my girlfriend Sara Khorshid for their love, their incredible support, their trust in me and for making all this possible. No matter what happens, I can always count on them. This work would not have been possible without them.

Gerald Bauer
Kaiserslautern, 2014

Contents

1

Motivation

"Over the next twenty years computers will inhabit the most trivial things: clothes labels (to track washing), coffee cups (to alert cleaning staff to moldy cups), light switches (to save energy if no one is in the room), and pencils (to digitize everything we draw). In such a world, we must dwell with computers, not just interact with them."

Mark Weiser, in [Wei96]
(1996)

More than 15 years ago, the vision of the confluence between the digital and the real was already envisioned by people like Mark Weiser. Since that time pervasive computing and the way people interact with technology has matured a lot. In the early 80's personal computers were barely found in common households. Today on the other hand, it is common for people to own multiple computing devices. This change was again strongly driven by the evolution of mobile phones to smartphones and the increasing popularity of tablet PCs which started about eight years ago. When walking through the streets one can see people of all age groups ranging from teenagers to the elderly having their mobile computers with them and using them in daily life situations. To be connected with people from all over the world through social networks such as Facebook and Twitter, to get location-based information adapted to one's personal interests or to have favorite city maps, photos and music in one's pocket at all times and the fact that these functionalities are integrated into one single lightweight device, was still a dream for most people several years ago. So what was the reason for this immense evolution? Of course, significant technological achievements have played their part. During the last few years sensors have been integrated into everyday items such as mobile phones. Embedded systems became more and more powerful and also smaller at the same time. These facts paved the way for complex data processing leading to innovative end-user services. Nonetheless, the best technology and the most assistive services would be useless if they were not accepted by society. Fortunately however, they are! Quite certainly two of the main reasons for this acceptance within all age groups are that innovative systems have been *unobtrusively* integrated into everyday life devices and that they have been designed to be usable *out-of-the-box* by technical laymen. So through mobile computing devices, the confluence between the digital and the real has already started.

Yet, the idea of "Ubiquitous Computing", which was introduced by Mark Weiser in 1988, is not only restricted to smartphones and tablet PCs. What about all the other everyday life objects? Have they also become smart during the past years? At least big companies using assembly lines or dealing with high security and safety issues are already using smart systems to reduce production time, to improve job safety or to save energy. For example smart working areas have been built to recognize what type of working step has been performed, to make workers aware of potentially dangerous situations in good time or to support them in order to reach a higher quality of service. Another example is related to pervasive health care systems where smart robots are used to support medical staff in high-precision and spatially restricted surgeries (e.g. cerebral tumor treatment). So-called living labs as operated by MIT or "The European

Network of Living Labs" group[1] (including more than 300 smart environments from all over the world) show that assistive services and related technologies are available for common households as well. In such environments smart services support people during daily life activities such as cooking, home control or energy saving. Examples are intelligent fridges that are able to order new food when necessary or systems that allow energy providers to overcome bottlenecks by controlling household appliances with high power consumptions – such as washing machines or laundry dryers – remotely. Looking at the above arguments, it becomes very clear that the confluence between the digital and the real has indeed already started. Be that as it may, when considering Mark Weiser's vision that the most trivial things in our environment will be smart, we can see that this is definitely not the case yet.

While some public buildings are at least using bus systems to control lights, shutters or heating, common houses usually do not have integrated smart systems (not to mention more complex assistive systems). As already noted, people of all age groups are enthusiastic about using mobile phone apps to support them during their daily life. But when it comes to smart applications for households the market seems to be nonexistent. RWE[2], HomeMatic[3] and NEST[4] are some of the few providers offering a complete smart home system. Yet, even these companies limited their portfolio to basic functionalities such as checking door and window status or heating control services. So what could be the reason for this? What distinguishes smart services on mobile phones from smart services integrated in homes? The most obvious reasons are related to one major point: *Applicability*. Whereas smartphones are unobtrusive devices which can be operated in an intuitive way out-of-the-box and applications can be easily purchased from a central app store with just a few clicks, the situation is quite different in the case of smart homes. Many approaches are based on highly instrumented environments which makes it hard to install services unobtrusively in real-life environments on a large scale. Besides that, many systems have to be initialized and adapted to specific environments or to users by experts based on large training data sets. Another reason might be the existing price model. Innovative smartphone applications are designed for the mass market and consequently a typical application can be purchased at a rather low price. In contrast to that, smart home systems or even single sensors can cost several thousand euros and many people are neither able nor willing to spend such amounts of money. As a final point, it should be mentioned, that the integration of smart systems and their extensions is still complicated and to a large part even impossible due to proprietary protocols and closed architectures. Only if these issues are solved, will smart homes have a real chance to conquer the mass market.

1.1 Thesis Direction

This thesis works towards the solution of the afore mentioned problems. It introduces a collection of multi-modal approaches for important issues in pervasive computing applications. More specifically, the following topics will be considered:

- Indoor positioning
- Use-mode recognition of common household devices
- Spotting and recognition of subtle, barely distinguishable hand activities and object interactions
- Detection of stress-related behavioral changes
- Standardized integration of pervasive computing systems in smart home environments

[1] http://www.openlivinglabs.eu/ (last accessed on 2013/06/04)
[2] http://www.rwe-smarthome.de/ (last accessed on 2013/05/10)
[3] http://www.homematic.com/ (last accessed on 2013/09/13)
[4] https://nest.com/ (last accessed on 2013/09/13)

The work shown was mainly driven by the EU project MonAMI[5], which aimed for the realization of activity recognition services in real-life and large scale scenarios. A major question faced in the project was: How can activity and context recognition systems be realized with a technology that:

- can be realistically expected to be deployed across several hundred existing, often old and so far non-instrumented homes,
- can be expected to work with minimal maintenance effort over months and years.

As a consequence all problem solutions observed in this thesis are based on the following three aspects that are of high relevance for such application fields:

- Minimizing training data
- Minimally invasive instrumentation
- The use of low-cost and multi-modal sensor systems

The importance of these aspects and resulting challenges are discussed in detail before the related work section, a thesis overview and the main contributions of this work will be introduced in the following.

1.1.1 Multi-Modal Activity Recognition Systems

Context encompasses more than just the user's location, because other things of interest are also mobile and changing. Context includes lighting, noise level, network connectivity, communication costs, communication bandwidth, and even the social situation [...].

Schilit et al., in [SAW94] (1994)

As stated by Schilit et al. in 1994, context is more than just location. Consequently, many approaches dealing with activity and context recognition issues are based on multi-modal sensor systems fusing various kinds of information.

Thereby single sensor modalities and their fusion can have more or less influence on the problem solution considered. In some cases even a single sensor modality can be totally sufficient to fulfill the planned task reliably. A simple example is the problem of providing location-based messages. When using GPS and WiFi location techniques even raw sensor signals are enough to localize a person outdoors and to provide environment-dependent information. In contrast, the recognition of more difficult activity recognition problems (such as nutrition), which consist in many cases of multiple user actions, require complex data processing techniques based on several sensor modalities. Hence, the selection of appropriate sensors and their fusion depends very much on the task at hand. Several attempts have been made to group activity and context recognition problems. In [Bla11] Ulf Blanke (ETH Zurich, Switzerland) introduced an activity hierarchy which ranges from the posture of humans over gestures, simple activities, complex activities and episodes of life. Another classification which is used by Paul Lukowicz (DFKI Kaiserslautern, Germany) in many talks is based on four connected blocks. Items observed are parameters (e.g. location, posture), state (e.g. modes of locomotion), actions (e.g. replacing a machine part) and situation (e.g. meeting, party). Although defined groups and hierarchies differ amongst each other, they clearly try to visualize the same vision of having complex context recognition systems that can be solved on the basis of multiple basic and simpler recognition systems. To illustrate this fact, I would like to pick up the example of nutrition monitoring previously mentioned. In order to recognize nutrition events such as preparing food, it is necessary to process and to fuse a bundle of basic user activity information. Besides the user's location in the kitchen and information about how kitchen devices are being used, systems which are able

[5] http://www.monami.info/ (last accessed on 2013/05/10)

to recognize gestures such as cutting and stirring are important basic components. By fusing information derived from such basic systems, the considered high-level context recognition task can be solved.

The thesis follows this concept. Amongst other scientific findings, a detailed evaluation will show the significant impact of multi-modal sensor systems and sensor fusion approaches on the recognition quality of a complex activity spotting problem. Thereby, information derived from several basic and stand-alone sensor systems (location, mode of locomotion, hand motion, device usage, etc.) is combined in order to get context information. In this way the search space of the considered problem can be reduced and the lack of large training data sets can be compensated. It is shown that the introduced multi-modal recognition system is able to outperform a state-of-the-art inertial system using extensive training data. To illustrate this vision again, this work shows a three-layer complexity hierarchy in Figure 1.1. The top layer includes recognition issues related to a person's life. These problems are the most complex ones as they are based on a large amount of context information which can be found in the middle layer. There, considered topics are nutrition, sports or social behavior patterns. Of course such recognition tasks can be only solved by fusing several basic pieces of information about user activities like mode of locomotion or performed gestures. Such recognition problems are located in the lowest level of the proposed hierarchy.

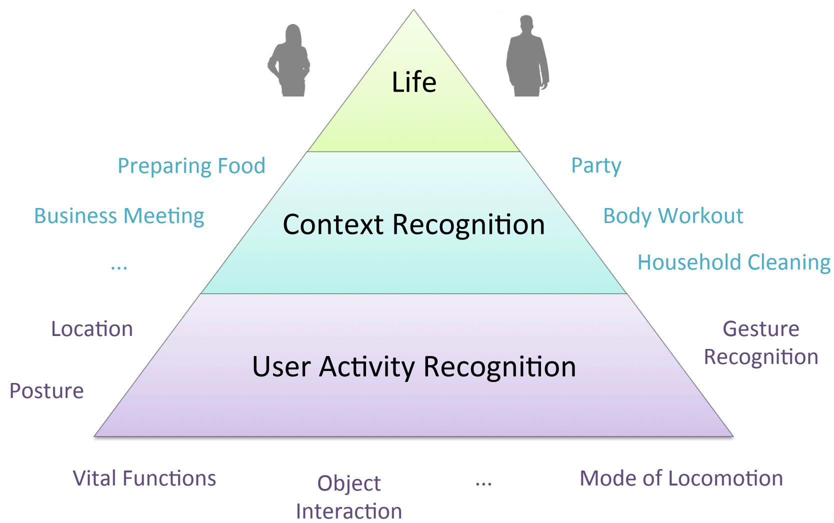


Figure 1.1: A three-level complexity hierarchy. Starting from basic user activity recognition systems, complex context recognition tasks and even life representations can be realized.

Another important aspect which depends strongly on the chosen sensor model is the maximum achievable recognition accuracy. A simple example is the issue of indoor positioning. While it is hard to accomplish positioning accuracies of a few centimeters when using low-cost sensor modalities such as RF based systems, expensive sensor systems such as the Ubisense (www.ubisense.net) or the Lukotronic (www.lukotronic.com) systems are able to solve the presented issue with ease (under specific conditions). Consequently, a trade-off between accuracy and cost has to be found in many cases – especially if the focus is on large scale and mass market applications. This work is limited to the use of low-cost sensor modalities and tries to overcome the loss of accuracy by multi-modal sensor fusion approaches. As a consequence, introduced solutions are usable in real-life scenarios on a large scale. Table 1.1 gives an overview about different sensor modalities that have been taken into account in this thesis.

Table 1.1: Considered sensor modalities.

Sensor Modality	Description	Cost
Cameras	Mainstream ceiling cameras as well as low-cost webcams have been used to track moving persons on a sub-room level and to identify objects.	600 / 80 €
Inertial Sensors	Inertial sensors have been used to recognize hand gestures, arm movement intensities and the hand position.	3000 €[A]
Microphones	Microphones which are built in mainstream Bluetooth headsets have been used to analyze sound samples coming from water pipes in order to roughly approximate the amount of water used.	< 50 €
Power Sensors	Power sensors have been developed and used to recognize how and what mainstream electronic household devices have been used for.	about 100 €
IR-Proximity Sensor	Proximity information has been used to realize the spotting of subtle arm actions and object interactions.	30 €
Smartphones	Built-in sensors such as GPS, Bluetooth and accelerometers have been used to realize mode of locomotion, room level location or high-level behavior recognition systems.	about 500 €

[A] Due to considerable facilitations with respect to integration, data acquisition and system setup, the Xsens system was used in this thesis instead of low-cost (below 100 €) alternatives.

1.1.2 The Challenge of using Minimal Amounts of Training Data

Context-aware computing is the ability of a mobile user's applications to discover and react to changes in the environment they are situated in.

Schilit et al., in [ST94] (1994)

In 1994 Schilit et al. mentioned that context-aware computing applications are able to discover and react to changes in the environment in which they are situated. In order to realize this ability, context and activity recognition systems have to learn and need to be adapted to both their environments and their users to a large part. The necessary training effort depends strongly on the chosen sensor system as well as on the focused recognition task. Consequently, the training effort could range from working out-of-the-box over simple configurations and recording minimal training data sets to collecting extensive training data sets. [Ogr09] focuses on a bicycle maintenance scenario, amongst other applications. There, 23 different activities should be detected by using nine inertial sensors and two ultrasonic beacons placed on the human body. The recognition system used is based on a training data set that includes 20 repetitions for each activity and for each person. In [Kun11] one of the evaluations shown is related to the Opportunity dataset (see [LPB+10]) including everyday life activities such as "making a sandwich" or "eating". The set of activities considered was performed by seven people wearing a sensor suit. Each person repeated the whole activity set five times. 33% of the collected data was used as a training set while the remaining data was used to evaluate the system. It is obvious that collecting such amounts of data is neither reasonable nor feasible in large-scale and real-life scenarios where a large number of activities and hundreds or even thousands of people are considered. If real-life pervasive computing systems are to be brought to the mass-market, they should ideally work out-of-the-box. However, this wishful thinking is currently only feasible for simple data processing tasks such as the recognition of locomotion modes. A well-known example is the commercial Nike+ portfolio[6] where acceleration sensors integrated in shoes are

[6] http://nikeplus.nike.com/plus/ (last accessed on 2013/05/10)

used to monitor sports activities.

Consequently, the challenge is to realize complex activity recognition systems which are based on minimal training data sets but are still able to provide reasonable results. The fact that the feature space considered may not be represented well when using too few training samples is a known issue in machine learning applications. Consequently, a tradeoff between the effort of collecting training data and the classification quality must be found. In this work, supervised learning techniques were focused on and only those training / configuration procedures were considered that are reasonable even for people with non technical background. Consequently, when talking about minimal training data sets, only data collection procedures based on simple and quick configurations or one-time measurements were taken into account. Especially when considering a large set of activities and multiple users, the importance of a reduced training data collection process will increase a lot.

1.1.3 The Importance of Minimally Invasive Instrumentations

The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it.

Mark Weiser, in [Wei99] (1999)

As stated by Mark Weiser, technology should be integrated into everyday life as unobtrusively as possible. It is obvious, that people are not willing to use systems that either don't suit their tastes or don't fit in with their personal surroundings. Besides, many elderly and especially people with dementia may feel very uncomfortable and uncertain when hundreds of sensors are placed visibly on objects or their clothes. As a consequence, the unobtrusive integration of sensor systems into real environments is of high importance if society's acceptance for those systems is to be achieved and if they are expected to make their way into everyday life. In general, systems can be characterized by the amount of object, user and environmental instrumentations needed. A dataset which was recorded within the EU project Opportunity clearly shows a highly instrumented setup. 19 on-body sensors, partly integrated in a sensor suit (see Figure 1.2, right image), were used to recognize user activities in daily life. Besides, even kitchen items were equipped with sensors as shown in Figure 1.2 (left image). Of course, the objective

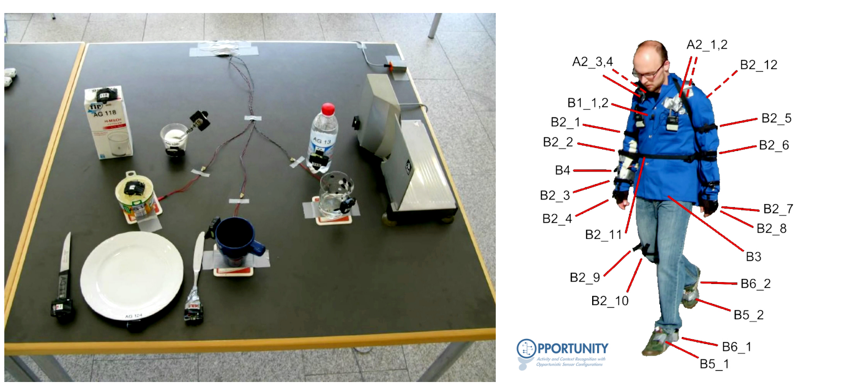


Figure 1.2: Highly instrumented environment. Pictures were taken from a promotional film of the EU project Opportunity (Source: http://www.opportunity-project.eu/challengeDataset; last accessed on 2013/09/15). Left: Instrumented everyday life objects. Right: On-body sensor placement (Each red line marks the location where sensors were placed to the body).

of this project was to give a proof of concept and not to provide a product which is close to the market. In [LHP+07] another highly instrumented environment is introduced. There, about 900 sensors were installed and used to detect daily life activities. On the other hand, systems were introduced that are completely integrated into everyday life objects as well. The most popular example is clearly the evolution of mobile phones to smartphones. Nowadays, smartphones invisibly integrate a set of sensors on which innovative services ranging from modes of locomotion detection (see [BL08]) to crowd estimation (see [JP11]) can be realized in a completely unobtrusive way.

This work used systems that can be unobtrusively integrated into existing environments. Sensors such as cameras, power meters and small microphones can be integrated subsequently and invisibly into real homes with ease. Besides, wearable on-body sensors can also be integrated into the user's clothes due to their small size and wireless communication. Finally, smartphones are one of the most unobtrusive sensor systems by nature.

However, the thesis does not focus on providing a ready-to-market product. Instead, it will give a proof of concept that solutions for relevant context and activity recognition problems can be found that fulfill requirements for real-life and large scale scenarios and are still able to provide excellent recognition rates. Consequently, the systems shown contain much more commercial potential than approaches such as the Opportunity system described in [LPB+10]. Figure 1.3 shows the thesis scope and a rough comparison with approaches that have been mentioned so far.

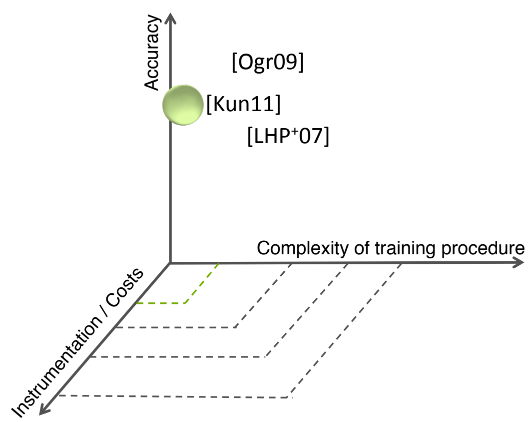


Figure 1.3: The thesis introduces various ubiquitous systems (represented by the green globe) that are based on low-cost sensors, low instrumentation as well as minimal training data sets and are still able to reach excellent recognition rates. Additionally, a rough comparison with approaches, that have been mentioned so far, is shown.

1.2 Related Work

This section discusses the current state-of-the-art in context recognition. Due to the fact that this thesis handles several different topics, a detailed topic-related state-of-the-art discussion is given in each chapter. A general overview about context recognition systems grouped by application fields is given in the following. Afterwards, a selection of state-of-the-art approaches is discussed with respect to the required training effort and environmental instrumentations.

1.2.1 Application Fields

In general, context recognition systems can be categorized by their field of application. Without a doubt, the most important and highly addressed fields are industrial applications, ambient assisted living, health care, sports, entertainment, social interactions and civil protection. A selection of publications related to these scenarios is discussed in the following.

1.2.1.1 Production and Maintenance

One of the most important application fields are industrial applications. The reason for this is obvious: Companies are very interested in integrating intelligent systems into their production or maintenance chains in order to reduce production time, increase quality of service and occupational safety or to reduce costs. Consequently, a lot of research has been done in this field. In [Ogr09] a car maintenance scenario is introduced. On-body ultra-wide band positioning systems as well as FSR units to measure muscular activity and a motion jacket including seven inertial sensors were used to reach the goal of identifying 20 common maintenance activities. Examples are opening/closing the hood and the doors of a car or writing on a check list. Besides car maintenance, the author also considered a bicycle maintenance scenario where 23 gestures related to wheel spinning, assembling and tightening/loosing screws were supposed to be recognized. Therefore, a reduced sensor set was chosen consisting of on-body ultrasonic positioning and inertial sensor systems. In [Bla11] and [ZWS09] the authors have picked up the car maintenance scenario again that was already shown in [Ogr09]. Thus, different approaches have been evaluated and compared with each other on the same dataset.

In [NSW+06] a wearable computing solution for aircraft maintenance scenarios is introduced. The approach shown does not focus on single technologies but on the combination of components with the objective to give a feasibility study. Aspects considered were the representation of and reasoning over maintenance data, user interfaces and clothes.

Several approaches were published related to furniture or wood workshops. In [Bla11] the construction of two wooden boxes was considered based on on-body acceleration sensors. In [LWJ+04] and [WLTS06b] 21 activities related to assembly tasks like "saw", "hammer" or "drill" were recognized by fusing on-body acceleration sensors with microphones. [AMS02] goes a step further and focuses on the determination of the current state of furniture assembly. The proposed system uses gyroscopes, force sensors and acceleration sensors on assembly items like a screwdriver or a hammer. The quality of the system was evaluated by a case study with the IKEA PAX wardrobe. Besides the recognition of basic gestures or actions, this work aims to provide a proactive guidance system. This aspect was also addressed in other publications like [Ten00].

1.2.1.2 Ambient Assisted Living and Healthcare

Ambient assisted living and healthcare scenarios are most certainly amongst the most important topics of the pervasive computing research community. Assistive services were mainly introduced in order to support people in their daily lives. Especially elderly and disabled people profit a lot from such systems as their quality of life can be improved significantly.

In [RSL12] a mobile platform is introduced that enables the monitoring of physical activities in daily life. The system is integrated into a healthcare system that is used to support out-of-hospital services. The topic of rehabilitation is addressed in [MJC+02] [BABMPBMA+12]. In [MJC+02] a wearable physical rehabilitation monitoring system is shown. The system consists of acceleration sensors and a PDA. It is used to generate real-time warnings in a hip and knee replacement rehabilitation scenario. A portable arm rehabilitation device is shown in [BABMPBMA+12] where a flex sensor, a force sensitive sensor and an accelerometer were used to realize a response system for rehabilitation activities. The monitoring of vital signs is used in [Par03] in order to help cardiopathic patients during rehabilitation. Another approach to monitor physical parameters was proposed in [CL08]. Using a capacitive sensor, activities such as breathing, breathing depth or drinking a cup of water can be distinguished. The topic of assistive systems in hospital scenarios was also addressed by many researchers. For example, in [CBL08] and [CBA+08] a capacitive sensor was integrated into the doctor's coat to provide a wearable user interface for documentary issues. Another unobtrusive way to get access to patient information in hospitals was introduced in [ABK+08] where a wrist worn acceleration sensor, RFID readers and PDAs were used to support doctors in their daily ward rounds.

Furthermore, the monitoring of nutrition is also a highly addressed topic in pervasive healthcare. Related approaches can be found for example in [AKT07b] [PSKL08] [CCC+09] [JCL]

[ASLT05] and [JGW04]. In [PSKL08] nutrition related activities like putting a piece of food into the mouth or drinking from a cup were considered. The proposed system is based on a magnetic field positioning technique. Using a single wrist-mounted sensor, the system was able to reach excellent classification rates (almost 100%) and also outperformed a standard inertial sensor approach. In [ASLT05] sound coming from the user's mouth is analyzed and used to detect what type of food has been eaten. A total of four different food types were distinguished and classification rates between 80% and 100% were achieved.

The topic of mental and physical health was addressed amongst other publications in [GOB+12] [JPS+13] [BBD+11] [RYNDL11] [MMPS10] and [PBLS10]. In [GOB+12] a real-life study was considered aiming at the monitoring of manic-depressive disorders. For this, smartphones and integrated sensor systems were used to determine features such as location, motion and phone call patterns. The system was evaluated by 10 patients and under the supervision of a psychiatric hospital. It is shown that the features considered are able to provide good indications for state transitions. The idea of using smartphones as sensors was also addressed in [BBD+11] and [RYNDL11]. The latter introduces MoodSense, which investigates the correlation between phone usage and the user's mood. The system was able to differentiate between four major mood types with an accuracy of 91%. In order to reach that accuracy, the system was trained using phone usage features collected during a period of three weeks. In [PBLS10] the PSYCHE system is introduced. It uses a monitoring system based on textile platforms and wearable sensors. Physiological parameters like electro-cardiograms (ECG), respiratory activity, galvanic skin resistance (GSR) and Electromyogram (EMG) were considered in order to identify signal trends that indicate the detection and prediction of critical events.

Besides the above, personal assistance approaches were addressed. In [LSB08] a system based on an electronic pen cover that is able to identify people by analyzing their handwriting was introduced. Additionally, the system is also able to perform text recognition and could achieve an accuracy of almost 80% on the IAM-OnDB-t1 benchmark task. In this way handwritten notes on paper or whiteboards can be stored digitally in real-time. In [TKSD11] an eye-tracking system working as an unobtrusive personal guide for museum visitors was introduced. The visitor's eye movements and detected eye fixations are synchronized with images delivered by a scene camera. Once a gaze on a specific object is detected, the object is recognized using computer vision techniques. Finally, audio files including detailed information about the recognized object are played back.

1.2.1.3 Sports

The recognition of sports and physical activities was also investigated in a lot of research. A large amount of publications deal with the recognition of modes of locomotion. In [FMT+99], [RM00], [VLC00], [LM02] and [MHS01] activities like walking, standing, running, walking upstairs and downstairs were considered. In [MHS01] two acceleration sensors were attached to a common belt. Based on processing algorithms such as PCA, motion activities like walking upstairs or downstairs were recognized. [VLC00] used acceleration sensors attached just above the knee and aimed to recognize activities such as jumping, climbing stairs or riding a bicycle. The scenario of riding a bicycle was analyzed in more detail in [KOL08]. There, acceleration and force sensors were used to recognize the gear in which the person is cycling or even the force applied to the pedals. Promising results up to almost 90% were reached in that scenario. Another sport scenario that was considered in a lot of research is martial arts (see [KBH+06], [HKG+06] and [PSKL08]). In [KBH+06] the focus was on recognizing Tai Chi gestures by using on-body motion sensors. Wearable acceleration sensors (eight sensors) were attached to the lower and upper arm, knees and neck. Two standard Tai Chi gestures ("Repulse the monkey" and "Parting the horses mane") were observed. The system was evaluated by using a 10-fold cross validation in combination with a kNN classifier. This resulted in a recognition rate of 76%. Furthermore, systems related to other sports scenarios like swimming (see [BFT09]) or baseball (see [LBG+09]) were also taken into account. Another interesting approach is shown in [WPVS05] where the IM4Sports (Interactive Music for Sports) system is introduced. The system consists of a wearable music player, a heart rate sensor belt, a pedometer using acceleration sensors and a wearable processing unit. The goal was to support the user during physical activity by adapting the music currently played to the needed motivation or the required training goal. In [MBL11] gym exercises were analyzed. The proposed system consists of a smartphone, that was worn in an arm holster. By analyzing motion data, 10 typical gym exercises like butterfly presses, chest press or arm curls should be spotted and identified. Besides, the amount of repetitions should be counted. The system was able to provide promising results for both tasks. In [WSP+11] the topic of paragliding was addressed. Flight information from paraglider pilots collected from their mobile phones was aggregated. It is shown that a real-time detection of collective behavior patterns leads to the discovery of regions with ideal thermal characteristics. Finally, even a skiing scenario was addressed in [HWT11] where the focus was on the topic of collision avoidance on ski slopes.

1.2.1.4 Entertainment

Entertainment is one of most profitable application fields. Home entertainment systems were revolutionized by the Nintendo Wii console in 2006. The new way of controlling games by hand motions introduced a completely new experience and game immersion. In 2010 Sony (Playstation Move[7]) and Microsoft (XBOX[8]) followed this trend and provided a similar and improved way of game control – both based on different sensor modalities. Even wearable gaming systems like the Playstation Portable[9] or Nintendo 3DS[10] and of course smartphones have integrated several sensor modalities and processing algorithms in order to realize new innovative games. Besides commercial products, the pervasive research community has also focused on approaches related to entertainment. [BAK+07], for example, introduced a parking game where a virtual car is controlled by hand gestures.

1.2.1.5 Social Interactions

The detection and analysis of social interactions and social relations was also observed in a lot of research. For example, in [PH08] and [Ogr09] office scenarios were addressed. Amongst other things [PH08] tried to recognize the fact if someone is answering the phone by using a glove equipped with 22 sensor units (CyberGlove II). [Ogr09] focuses on 16 activities related to the topic of "giving a talk" using a FSR glove and inertial sensors. In [YYA+10] a multi-modal approach using microphones, cameras and motion sensors was shown aiming at the acquisition, recognition and visualization of human interactions within a meeting scenario. This scenario was also addressed by many other approaches as in the Conference Assistant [DSAF99], Easy Meeting [CFJ04] or Team Space [RAG+01].

In [EPL09] observational data from mobile phones were analyzed to infer friendships amongst other things. In [API+11] the "Friends and Family" study is introduced. There, a family was transferred to a living laboratory and the aim was to measure, to understand and to design social mechanisms in the real world. Finally, in [ZDC11] a socioscope model for social network and human behavior analysis based on information about mobile phone call details was introduced. The objective of this approach was to quantify social groups, relationships and communication patterns and to detect human behavioral changes. The system was evaluated by using real-life call logs from 81 users that were collected over an eight month period. The resulting set contains more than 500.000 hours of data on the users' location, communication and device-usage behavior.

1.2.1.6 Civil Protection

The issue of public security and civil protection gained a lot in importance over the last few years. Most certainly, the mass panic that took place at the Loveparade music festival in

[7] http://de.playstation.com/psmove/ (last accessed on 2013/09/15)
[8] http://www.xbox.com/ (last accessed on 2013/09/15)
[9] http://de.playstation.com/psp/ (last accessed on 2013/09/15)
[10] http://www.nintendo.de/Nintendo-3DS/Nintendo-3DS-94560.html (last accessed on 2013/09/15)

2010 has proved once again the increased need for reliable panic detection systems. For many years research approaches were presented related to critical life-saving scenarios such as rescue operations.

In [FZ09] a vibro tactile directional guidance is introduced. The system consists of a wearable belt-like device that notifies individuals about exits in panic situations. Simulations have shown that the system is able to slow down panic growth. Besides, the number of evacuated people could be increased. New navigation approaches were also introduced in [KRG+] and [Kla08]. The proposed lifeNet system consists of an ad-hoc sensor network providing a relative person localization. In combination with a wearable firefighter support system, an indoor-navigation system for reduced visibility scenarios was realized. Firefighter scenarios were also addressed in [Kla07] and [JCH+04] where a system is introduced that supports the communication between firefighters and provides the basis for gathering, integrating, and distributing contextual data, such as location and temperature.

Beside systems that support safety or emergency forces, a new research field has been rising during the last few years: The analysis of crowd behavior and crowd monitoring. The objective of these systems is to detect danger at people gatherings in time and to prevent mass panic. In [WFMK+12], [WFR+12] and [RWHT11] mobile sensors were used to detect the emergence of crowds. In [WFR+12] mathematical methods are introduced to infer and visualize crowd density, crowd turbulence, crowd velocity and crowd pressure in real time. In order to reach that goal, pedestrians' GPS location traces were analyzed. The approach was tested during the 2011 Lord Mayor's Show in London where around four million location updates from over 800 visitors were analyzed. Another approach to estimate crowd density is based on Bluetooth device discovery. In [JP11] mobile phones are used to scan for nearby Bluetooth devices which belong to other people. A three day experiment at the Oktoberfest in Munich showed that such a system is able to recognize four discrete crowd density classes with an accuracy of over 80%. Similar approaches are shown in [NK07b] and [MBC09].

1.2.2 Related Work, Training Complexity and Environmental Instrumentation Issues

The last section showed that a large amount of approaches focusing on the solution of real-life issues have been addressed by the pervasive computing research community. The systems introduced were able to solve the problems observed with reasonable or even excellent recognition accuracies. But the most important question remains: Are these systems also usable in real-life and large scale scenarios? As already mentioned in previous sections of this thesis, minimal training data sets, minimally invasive environmental instrumentations and the usage of low-cost sensor modalities are essential for this field of application. Hence, in the following activity and context recognition approaches are discussed in more detail with respect to these aspects.

In [Ogr09] a real-life bicycle maintenance scenario is introduced. The proposed system aims at spotting and recognizing 23 different gestures related to the considered scenario like "pumping", "tightening/loosing screws" and "turning pedals". The system consists of wrist worn sensors: Xsens accelerometers (1000 € per sensor) and Ultrasonic devices (Hexamite HX900, 1300 € for a basic setup). In order to track the person's hand, four Ultrasonic base stations were integrated into the environment. Neglecting the necessary effort to adapt the Ultrasonic system to a specific environment (which quite certainly can't be done by technical laymen), the feasibility of a real-life and large-scale deployment of this system is ruined by the required training effort. 20 repetitions for each of the considered gestures were used to train the system. In total, 3035 instances were collected from multiple people which results in 291 minutes of training data. Besides bicycle maintenance the same work also addresses the topic of car assembling. 20 gestures such as "open/close a car's hood", "check trunk gaps" or "lock check left" were focused on. The UBISENSE[11] ultra-wide-band (UWB) user positioning system was one of the main components of the proposed system. The fact that this system costs more than

[11] http://de.ubisense.net/en/ (last accessed on 2013/07/22)

15.000 € makes the presented approach unaffordable for the mass market. Even the increased amount of necessary on-body sensors (two FSR sleeves to monitor muscle activity and a motion sensor jacket including more than nine inertial sensors) destroys the idea of an unobtrusive system. Finally, the author mentioned that people were instructed about how activities have to be carried out in both scenarios. This artificial user behavior is obviously far from real-life conditions. However, the fact that [Bla11] and [ZWS09] have also focused on the same dataset confirms the importance of this scenario.

In [Bla11] body-worn inertial sensors were used for various real-life application scenarios. Besides the introduced car quality check scenario, the "Drink and Work" (spotting several drinking events) and "Woodshop" (building a wooden bookshelf) datasets were considered. The systems shown aimed at recognizing a total of 42 activities which were carried out multiple times by different users. To reach a user independent system, a leave-one-user-out cross-validation was performed. However, the fact that the training procedure also requires information about background activities (the training set includes 20% of recorded background data), makes the system hard to deploy in large-scale and real-life applications. In such scenarios, it is nearly impossible to collect data sets covering all feasible background activities in a reasonable time. The resulting problem of having an incomplete background data set and its negative influence on the recognition quality was also mentioned by the author himself.

In [Kun11] the "Drink and Work" and the bicycle maintenance scenarios were picked up again. Additionally, a data set containing daily life activities like "ironing", "packing" and "washing" within a home scenario as well as the Opportunity dataset already mentioned in Section 1.1.3 were considered. The work shown is based on on-body acceleration sensors and uses Hidden Markov Models to recognize user activity. The training data set includes 33% of all recorded data, which makes it again hard to deploy on a large scale and in real-life scenarios.

In [AGS12] a hands-free interaction system based on 3D handwriting recognition is shown. Users can write text in the air as if they were using an imaginary blackboard. A wireless data glove including an acceleration and gyroscope sensor was used. The system is able to achieve a word error rate of 11% in a user independent case. In a segmentation step the system tries to separate segments containing handwriting activities from background data. The training data set used includes 383 minutes of recorded data. 111 minutes thereof are background activities. As already mentioned before, the need for background training data makes it nearly impossible to deploy the system on a large scale in various application fields.

In [RYNDL11] smartphones were used to infer the owner's mood. All in all 25 iPhone users were studied and correlations between their mood and their phone usage were derived. Based on user independent smartphone usage statistics like the amount of phone calls, application usage and location, four mood types could be recognized with an average recognition accuracy of 61%. When considering user-dependent features, the recognition quality could be increased to 91%. However, the system shown needs an immense amount of training data. All in all three weeks of collected phone usage data were used to train the system. It is obvious that behavior recognition systems are in general more complex than systems related to basic user activities. As a consequence much more training data is needed to set up such systems. Once again, this fact makes it hard to deploy the system in real-life and large-scale scenarios and to motivate consumers to use the system.

[PSKL08] introduced a wearable system related to two real-life applications: The recognition of four Tai Chi gestures (e.g. "Parting Wild Horse Mane" and "Single Wip") and six daily life activities like "touching the skin" or "drinking from a cup". A novel relative positioning technology based on magnetic field measurements was applied. It turns out that the system is able to outperform a state-of-the-art inertial approach and achieves outstanding recognition rates between 94% and almost 100%. However, the system needs a significant amount of training data. In the case of the Tai Chi scenario, the system was trained by using 33% of the recorded data. This number was exceeded by adapting the system to recognize daily life activities. In that case, even 50% of all recorded data (6500 training samples) were used to train the system.

Many computer vision based approaches are dependent on large amounts of training images. The INRIA dataset is a widely used and referenced data collection for human detection applications. The provided set contains more than 1800 training and about 740 test images.

Publications like [WHY09], [DT05] and [Dal06] have trained and evaluated their proposed solutions on this dataset. Besides, the MIT pedestrian data set consisting of 709 pedestrian images is also a well known benchmark collection. It is used by many publications like [POP98] [PP00] and [MPP01]. In [Dal06] the MIT dataset was split into two sets: a training set containing 509 images and a test set of 200 images. SIFT based techniques [SRE+05] [PWF09] were used in many object recognition applications. SIFT is invariant to rotations, scale changes and illumination changes. However, they are based on an immense number of key points which has limited its usage in object recognition applications. Systems shown in [PWF09] and [SRE+05] were benchmarked on three well-known datasets: Caltech (101 categories; about 40 to 800 images per category; the recommended ratio between training and test images is between 1:30 and 1:1), Coil-100 (7200 images from 100 objects such as hamburgers, various cups or cars) and an MIT dataset (including 2873 images of indoor and outdoor scenes). In [ZBMM06] the proposed system was evaluated on several datasets. One example is the MINST dataset which consists of handwritten digits and contains 60.000 training images and 10.000 test images. Besides, the USPS dataset includes 9298 handwritten digits of which 7291 are used for training and 2007 for testing. Finally, the CUReT dataset was introduced that includes 61 images of real world textures. There, 50% of the included images were used for training.

Besides the approaches introduced so far, a substantial amount of activity and context recognition systems are based on statistically relevant and large training data sets. Further examples are shown in [KBH+06] [ASLT05] and [MBL11].

During the last few years more and more living lab environments have been constructed. A living lab is a real-world environment in which sensors and observational technologies are integrated and used to benchmark context and activity recognition systems. At this point I want to once again pick up the Opportunity dataset which is described in [LPB+10]. As already mentioned, the dataset consists of natural human activities (e.g. "waking up", "grooming" or "cleaning") which were recorded in a sensor rich living lab environment. The data was recorded by on-body systems as well as sensors that were integrated into the environment or were attached to daily life objects (e.g. bread cutter, spoon or plates). All in all 15 networked sensor systems (proprietary and custom; from different providers and research labs) were considered including 72 sensors and 10 modalities. It is clear, that such highly instrumented environments are far from real-life large-scale applications. This rather invasive scenario is topped by [LHP+07] where about 900 sensors were deployed in a home environment. Wired reed switches, current and water flow monitoring systems, RFID tags and motion detectors were installed. Daily life activities like "reading" and "eating" were addressed by this work. The authors highlighted the fact that data was recorded under real world conditions. However, a large scale deployment is clearly hard to realize due to the enormous amount of installed sensors.

A highly instrumented and one of the most famous living lab environments is PlaceLAB [ILT+06] [ILB+05]. PlaceLab was opened 2004 in a urban neighborhood in Cambridge, Massachusetts. The 1000 sq. ft. apartment consists of several rooms and a large amount of integrated sensors. For example 15 cabinetry components each containing a micro controller, speaker systems and sensor networks of up to 30 devices (recording of audio-visual activity) were created. Wired sensors were unobtrusively integrated into the cabinetry, appliances and furnishings whereas small wireless sensors were placed inconspicuously on objects. Apart from that, 80 wired on-off/open-close switches, 34 temperature sensors, 10 humidity sensors and 37 electrical current sensors were integrated to name just a few. All in all, more than 350 sensors were deployed until December 2005. Figure 1.4 shows an example of sensors that were integrated in a cabinetry.

Another living lab was introduced by the Georgia Institute of Technology. The Aware Home [KOA+99] is a living laboratory for ubiquitous computing applications in daily life scenarios. Sensor systems such as smart floors and RFID tags attached to daily life objects were considered. In [ABE+] the Aware Home environment was used to develop and test technologies for successful aging.

The TUM kitchen dataset [TBB09] introduces another living lab environment. The intelligent kitchen aims at realizing assistive technologies in order to support elderly or disabled people and to improve their quality of life. Four static cameras were used to capture human

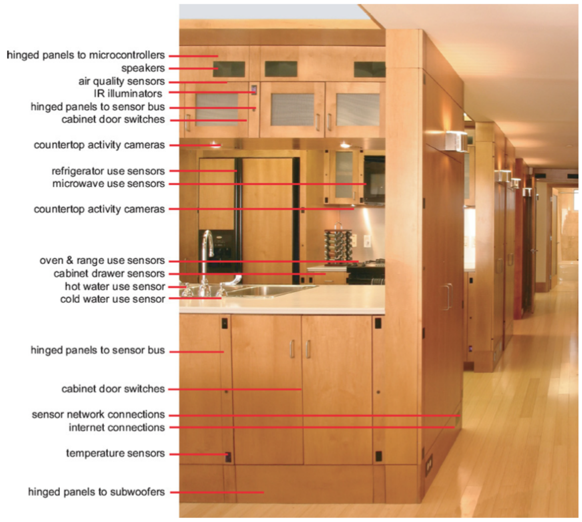


Figure 1.4: PlaceLab: Sensors integrated into a cabinetry (Source: [ILT+06]).

motion. Furthermore, magnetic sensors were attached to doors and drawers in order to detect opening/closing. Finally, three static readers were integrated into the environment and RFID tags were attached to items like napkins, cups or knives.

In [ZÖ9] [MPK10] SmartFactory is introduced. It is a demonstration and research test environment for future factory technologies which belongs to the German Research Center for Artificial Intelligence (DFKI) in Kaiserslautern, Germany. Common factory components have been equipped with various sensor modalities such as Bluetooth, ZigBee, UWB, NFC and RFID systems. The Innovative Retail Laboratory (IRL) [KSJ10] is also a research laboratory of the German Research Center for Artificial Intelligence (DFKI) and is operated in collaboration with the German retailer GLOBUS SB-Warenhaus Holding in St. Wendel. The IRL includes a variety of smart items with digital product memories (a further development of the RFID technology) as well as indoor positioning and navigation systems. The main objective is to evaluate new ways of customer interaction in a real world environment.

Besides the previously discussed living labs and smart home environments, a large number of similar intelligent and instrumented environments exist all over the world. In [BL10] a detailed overview of well-known living labs and an estimation of the level of included intelligence is given. Finally, more than 300 international smart environments are listed in the "European Network of Living Labs" group[12].

In summary, it can be seen that the pervasive research community has introduced an incredible amount of activity and context recognition systems as well as living labs and smart environments in general. The overall vision is clear: the proposed systems have been designed to solve problems related to real-world problems and to be deployed in real-world environments. However, it could be seen that a large part of these systems have not taken the aspect of applicability into account which includes the need for minimal training data and minimal invasive instrumentations. Nevertheless, the importance of these requirements has already been addressed by some approaches. In [Bla11] a way to significantly reduce the required amount of training data is introduced by transferring learned knowledge from one application to others. Another approach, which is shown in [DBL11], introduces a way to integrate new sensors into an existing system without re-training it. This fact also leads to a reduced training set. It is obvious that in terms of image recognition systems the described amount of necessary training images can't be collected in large-scale scenarios where a large number of different objects is considered. To overcome this problem some computer vision approaches are using so-called one-

[12] http://www.openlivinglabs.eu/ (last accessed on 2013/06/04)

shot-training methods. The objective is to significantly reduce the amount of training images. In [FFFP03] a system is introduced that is trained by only one to five images. The shown system uses a Bayesian one-shot learning algorithm which is based on "generic" knowledge which may be obtained from previously learned models of unrelated categories. The system was evaluated on datasets containing several categories (like faces, airplanes and motorbikes) and could achieve comparable results to exiting approaches. Another one-shot approach is presented in [FFFP06]. There, object categories were considered and it is shown that it is possible to learn information about a category from just one or a few images. The system was evaluated on a dataset containing 101 diverse object categories. The idea of an one-shot computer vision approach that uses a single image to identify objects is also considered in this thesis in Chapter 4.

A large amount of approaches use semi-supervised learning instead of supervised learning techniques (see [CSZ06]). In this way the amount of labeled training data needed can be reduced. Systems based on concepts such as self-training (e.g. [CSZ06]) and co-training (e.g. [BM98]) have been already used in many activity recognition scenarios (e.g. [SVLS08] [GYL+07]). Further examples and a more detailed discussion about this topic can be found later in Chapter 4.

This thesis introduces various solutions for activity and context recognition systems by taking requirements of real-world, large-scale applications into account. Table 1.2 again shows the scope of the thesis and elected related approaches.

Table 1.2: The scope of the thesis and elected related approaches in terms of minimal training data sets, minimal invasive instrumentations and affordable sensor systems (Personal assessment – ranging from "+" (not fulfilled) till "++++" (completely fulfilled). "-" indicates that a rating cannot be given.

Approach	Minimal training data	Minimal invasive instrumentation	Affordable sensor systems
Thesis	+++	+++	+++
[Ogr09] (Bicycle scenario)	+	+	+
[Bla11] (Woodshop scenario)	+	++	+++
[LPB+10] (Dataset)	-	+	+
[LHP+07] (Dataset)	-	+	++
[RYNDL11]	+	+++	+++
[PSKL08]	+	++	++
[FFFP03]	+++	+++	+++

Thesis and [FFFP03]: Using low-cost accelerometers instead of Xsens inertial sensors.

1.3 Thesis Outline

The following section shows the overall structure of this thesis. As shown in Figure 1.5, the thesis covers two issues:

- Solutions for activity and context recognition problems using minimal training data, minimal environmental instrumentations and low-cost multi-modal sensor systems (main part)
- Standardized integration of pervasive computing systems into smart home environments (closing topic)

I want to highlight again, that the main focus of the thesis is clearly on the first issue which will be discussed in Chapters 2 to 5. The thesis is complemented by the proposed standardization approach in Chapter 6, which will gain in importance if context recognition systems related to real-life scenarios are to be realized in real smart home environments on a large scale. In the

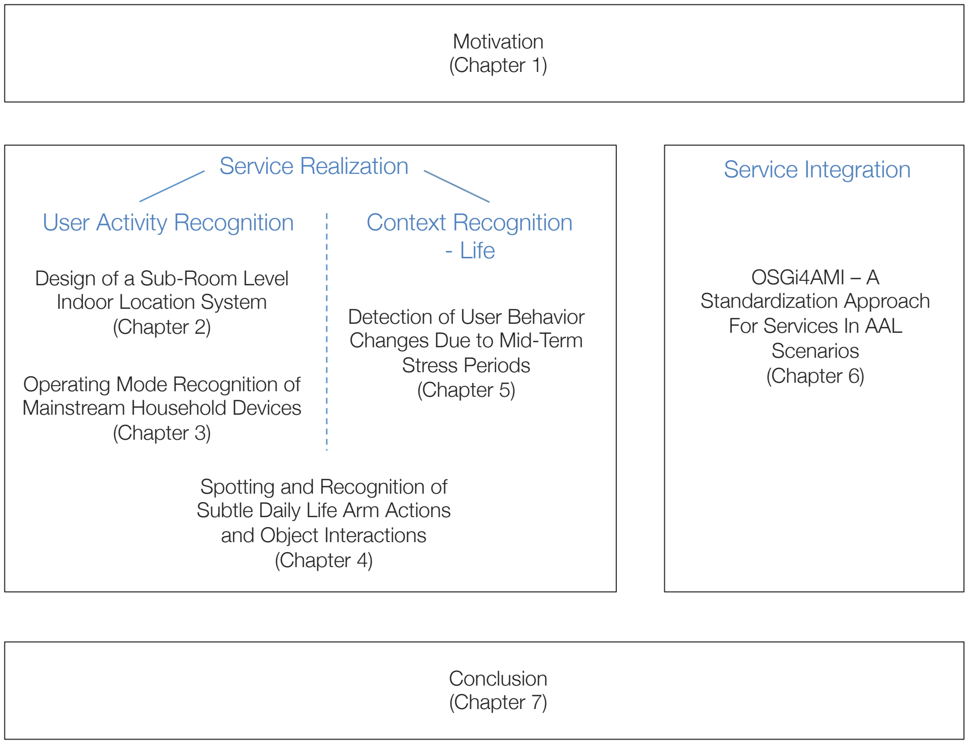


Figure 1.5: Thesis Chapters.

following, each chapter is only briefly outlined as detailed summaries and thesis contributions can be found in Section 1.4.

Chapter 2 - Design of a Sub-Room Level Indoor Location System. This chapter addresses the issue of indoor positioning. The approach shown is able to track and identify people on a sub-room level. Consequently, location information related to pre-defined regions of interest (e.g. bed, sink, fridge or sofa) can be provided. The system is realized based on the correlation of unidentified moving objects derived from a standard and simple computer vision tracking system (providing a reliable sub-room level positioning) with on-body acceleration sensors (solving identification issues) through motion pattern analysis. The system is evaluated by focusing on a scenario that includes two people performing different tasks within the same room at the same time. This chapter is based on my work published in [BL08].

Chapter 3 - Operating Mode Recognition of Mainstream Household Devices. Besides the location of people, the current operating-mode of household devices is a valuable source of information for context recognition applications. So far, common household appliances are largely unable to provide information on how and what they have been used for. This chapter introduces two novel approaches to turn common household appliances into smart devices. First, the status of single water taps is detected and water consumption is approximated through sound analysis by an off-the-shelf low-cost Bluetooth microphone. Second, operating modes of common electronic devices and information about *what* these devices have been used *for*, is recognized based on an electric current sensor (*iSensor*). Single or even multiple devices (using multi-plugs) can be operated by the *iSensor*. Based on rules and power consumption profiles, electric devices and their use-modes can be recognized. Both systems were evaluated in several real-life scenarios and by multiple users. I want to note, that the design of both sensors was a joint work with Alejandro Ibarz (Tecnodiscap, University of Zaragoza, Spain) and Karl Stockinger (Embedded Systems Lab, University of Passau, Germany). Although much research work has been done on similar or related topics during the last few years, I want to emphasize, that the systems introduced here can be seen as pioneer work and are amongst the first publications in this research field. This chapter is based on my work published in [IBC+08] and [BSL09].

Chapter 4 - Spotting and Recognition of Subtle Daily Life Arm Activity and Object Interaction. This chapter focuses on the spotting and recognition of subtle arm activity and corresponding object interaction. The activities observed were embedded in a continuous data stream containing a large amount of daily life background actions. Object interactions have been spotted and identified using a core on-body system. The system is able to provide reasonable recognition accuracy and is therefore used as a starting point for further fusion approaches. Stand-alone sensor systems (amongst other approaches, solutions and concepts from Chapter 2 and Chapter 3 are being picked up again) are used to reduce the considered search space (in order to compensate the lack of minimal training data) and are fused step-by-step with the core system. Moreover, several extensions and improvement approaches are shown for the core system and evaluated in-depth. Finally, subtle hand activities are recognized based on the object interactions detected. Therefore, three different fusion concepts are introduced and compared with a state-of-the-art inertial approach. All approaches were evaluated within an office environment and under conditions close to real-life. This chapter is based on my work published in [BBLS13].

Chapter 5 - Detection of User Behavioral Changes due to Medium-Term Stress Periods. So far, solutions for well-known activity recognition problems have been introduced. This section shows how complex behavior patterns of individuals and deviations due to medium-term stress periods can be detected by smartphones and integrated sensor systems. The system introduces several features related to the user's location, social interactions and mobile communication patterns used to describe the behavior of people. A 24-7-4 (24 hours, 7 days a week, 4 weeks) real-life data recording was performed with multiple users. The objective was to evaluate chosen behavioral parameters with respect to their ability to represent stress related behavioral changes. This chapter is based on my work published in [BL12].

Chapter 6 - OSGi4AMI - A Standardization Approach for Services in AAL Scenarios. Chapter 6 complements the thesis by introducing a standardization approach for activity and context recognition problems related to real-life ambient intelligence applications. This way pervasive computing systems as introduced in this work can be easily integrated in smart home gateways. Moreover, the resulting interfaces enable a distributed service development and the re-usability of existing services in smart home environments. Concepts and systems that were introduced in Chapter 2 and Chapter 3 are used to illustrate the standardization process. Finally, the development of end-user services is described based on distributed implementations of basic services and defined interfaces. These services were deployed in several real homes and were used by residents for about two months. This chapter is based on my work published in [MCB+09] and [FKW+10].

Chapter 7 - Conclusion. This chapter summarizes the systems introduced and the achieved results presented in this thesis. Additionally, open issues related to smart home systems and large-scale real-world scenarios, which were solved by the EU project MonAMI, are discussed.

Finally a general outlook is given.

1.4 Contributions

The main objective and overall contribution of this thesis is to provide reasonable solutions for activity and context recognition problems and at the same time

- *to minimize the amount of training data (simple one-time measurements or data collections that can even be performed by technical laymen in a reasonable time),*
- *to minimize intrusive environmental instrumentations (unobtrusive environment integration or on-body placement),*
- *to use multi-modal low-cost sensor systems (sensor systems must be affordable for average earners) and*
- *to perform evaluations under real-life or close to real-life conditions*

In this way the realization of activity recognition systems in real world environments on a large scale is pushed on. Figure 1.6 visualizes the overall structure of the thesis again. In the following, research questions and the main contributions are shown separately for the main (solutions for activity and context recognition systems) and closing part (smart home integration) of the thesis.

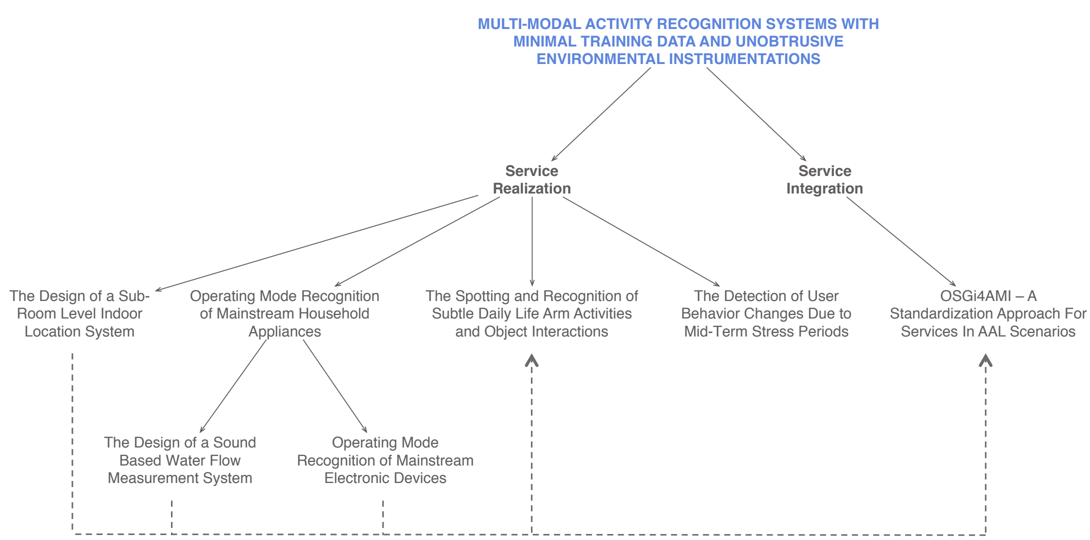


Figure 1.6: Overall thesis structure.

1.4.1 Solutions for Activity and Context Recognition Systems

In Chapter 2 a solution for localizing people indoors on a sub-room level is proposed. The system is based on standard and simple computer vision techniques able to track unidentified objects on a sub-room level using ceiling mounted fish-eye cameras. Thus, the main research questions were:

- *How can a standard and simple computer vision-based tracking system be enriched with information about a person's identity by fusing it with on-body sensors?*

- *Are motion pattern correlations a suitable solution for the considered problem?*

- *Can smartphones and their built-in motion sensors be used in this context to guarantee an unobtrusive system realization?*

- *Can the system be realized under the condition that privacy issues which occur when sending personal data to a central processing unit have been overcome?*

The main contribution of this chapter is that a simple computer vision sub-room level localization approach can be enriched with information about the user's identity. This is realized by correlating motion patterns derived from user assigned on-body motion sensors and vision tracked unidentified persons. A feasibility study showed that the system is able to provide a sub-room level positioning of individuals almost in real time (time deviation of a few seconds) under close to real-life conditions. The big advantages of the proposed system are that it works nearly out-of-the box (just simple one-time configurations have to be performed) and due to the fact that it is based on smartphone integrated accelerometers and ceiling cameras, the system is completely unobtrusive and affordable. Besides, the system can be realized in a way that neither personal data nor images have to be transmitted. Only information about motion patterns and sub-room level positions of unidentified moving objects are sent to people's mobile phones. Hence, privacy concerns are weakened. Finally, another very important advantage of the system is, that it is easily scalable as new smartphones have just to register with the system in order to receive data about unidentified moving objects. This work was cited in 13 publications[13] during the last few years.

Chapter 3 introduces novel ways to turn mainstream household appliances into smart devices. The systems described are based on low-cost microphones and power measurement sensors (*iSensor*). The intended objective was to approximate the rough amount of water consumed from single water taps and to recognize operating modes of electronic devices as well as what they have been used for. The resulting research questions were:

- *How can mainstream household appliances (water taps and electronic devices) be turned into smart devices (able to provide their operating mode and the reason what they have been used for) by people with basic technical knowledge?*

- *How well can the amount of water used be approximated by using minimal training data including reference sound samples of only a few different water flows? And how well can surrounding sound be filtered out without recording reference noise samples?*

- *Are simple power profile based rules able to recognize use-modes of electric devices and the reason what they have been used for? Can multiple devices be operated at the same time using a single iSensor?*

The main contribution of the water measurement system was to define and to combine methods and rule sets, so that the amount of water consumed can be approximated by using only a rather restricted amount of reference data. This implies, that surrounding noise, that usually appears in common kitchen and bathroom environments such as talking people, music or noise coming from surrounding appliances has to be filtered out without recording reference samples. Otherwise the idea of minimal training data would be destroyed due to the large amount of different noises that could appear. The proposed system is based on a low-cost Bluetooth headset. The microphone analyzes sound samples from water running through inflow pipes of water taps. Due to the sensor size and design, it can be integrated into existing environments

[13] Source: Harzing's Publish or Perish (last accessed on 2013/05/31)

completely unobtrusively. Various real-life and real-time evaluations (lasting up to 35 minutes) were performed in real kitchen and bathroom environments. Although the system was trained with a minimized data set that consists of only seven reference water flows, it delivers an average absolute error of only 10% during these experiments (all in all 76 liters of water were used).

With respect to the recognition of electric device use-modes, one of the main contributions was to design a model based system that uses only simple processing methods and simple threshold based classifiers. The reason for this was, that in this way data processing algorithms could be embedded in the *iSensor* itself and hence only high-level context information needs to be transmitted. As the resulting sensor is a stand-alone system, it is easy to deploy, to maintain and to extend. To solve the considered problem of turning electric appliances into smart devices, rule based classifiers in combination with power profiles and time features were developed. As a consequence such systems can be operated even by simple and low-cost micro-controllers as they have been used for the *iSensor*. The system was evaluated with eight mainstream kitchen appliances (e.g. water boiler, bread cutter or coffee machine) and by multiple users in several environments and under real-life conditions. The system showed very good recognition rates (84% and more) for all devices and besides the devices' operating modes (e.g. current intensity level of a food blender) the intention of the user could also be recognized (e.g. stirring something liquid until it gets creamy). Furthermore, the ability of the *iSensor* to handle several connected devices operated at the same time was evaluated.

Although many publications have been dealing with these topics during the last year, the proposed systems are amongst the first approaches in this field and can be seen as pioneering work. Both approaches have been referenced all in all by 31 citations[14]. Additionally, both systems have been awarded ("Best Commercial Potential Award" and "Best Paper Award").

In Chapter 4 the issue of spotting and recognizing subtle arm activities and corresponding object interactions within a continuous data stream is considered. Since only subtle and barely distinguishable activities lasting only a few seconds at best (e.g. pushing buttons on different devices) have been taken into account, they are hard to distinguish amongst themselves as well as from background activities. The difficulty of the targeted problem is also confirmed by the fact that a state-of-the-art and well elaborated inertial sensor approach designed and applied by Ulf Blanke (ETH Zurich, Switzerland) was not able to reach sufficient recognition rates. A core system that is based on on-body components (wrist camera, proximity sensor and inertial sensors) is introduced. Although the system does not need any infrastructure instrumentation at all and is trained by simple one-time measurements and recordings, it is able to provide reasonable recognition rates. This configuration was used as a starting point for further sensor fusion approaches. The following research questions were considered:

- *How can subtle and barely distinguishable arm actions and corresponding object interactions be spotted within a large set of daily-life background data considering the previously mentioned system requirements for large-scale and real-life scenarios (minimal training data, minimal intrusive instrumentations and low-cost sensors)?*

- *Can the achieved recognition quality of the core system be significantly increased by fusing it with additional stand-alone sensor systems? What is the impact of each additional sensor modality? How far can the search space of the considered problem be reduced by sensor fusion? Keeping in mind that all considered systems have to fulfill the mentioned requirements for real-life and large scale applications.*

- *What is the impact of integrating a state-of-the-art inertial sensor system using large amounts of training data?*

- *What is the impact of replacing the wearable camera of the core system by a state-of-the-art inertial system?*

[14]Source: Harzing's Publish or Perish (last accessed on 2013/05/31)

- *How well can subtle arm actions be recognized based on detected object interactions in combination with a trained inertial system, information from smart appliances and a combination of both?*

The key contribution of this chapter is the definition of model based approaches and systems based on minimal training data as well as the definition of appropriate fusion methods, that are able to solve the considered spotting problem and to significantly outperform a state-of-the-art motion based approach. Moreover, an in-depth evaluation of several multi-modal sensor fusion approaches in terms of their impact on the recognition quality of the core system is shown. Therefore, eight stand-alone systems based on simple one-time configurations and minimal training data sets related to indoor positioning (the basic idea of Chapter 2 is being picked up again), motion pattern analysis, time features and smart appliances (mainstream appliances were turned into smart devices by using systems introduced in Chapter 3) and their combinations were fused with the core system. The main idea was to restrict the corresponding search space by sensor fusion and in this way to compensate for the lack of minimal training data. Furthermore, the impact of the state-of-the-art inertial system (developed by Ulf Blanke) was evaluated as well. It was also analyzed, if the core system is able to achieve similar results when replacing its key component (a wrist mounted wearable camera) by a state-of-the-art inertial system. All in all 22 system fusion approaches were evaluated based on a data set that has been recorded by multiple users under close to real-life conditions and under the aspect of identifying object interactions. The best system combination was able to improve the EER of the basic system by 32% to 79%. Compared to the state-of-the-art inertial system the introduced system could raise the EER by a significant 61%. The best system configuration was chosen as a starting point to spot and to recognize underlying subtle hand movements. Therefore, three fusion approaches were introduced. The systems were able to raise the EER between 32% and 57% compared to the inertial sensor system. Depending on the fusion approach, hand activities could be recognized more or less precisely. As a consequence activity sets including 25, 27 and 32 activities were considered during the evaluation.

Chapter 5 focuses on the issue of recognizing user behavior and significant behavioral changes. In this context a system is introduced that defines behavioral user parameters including location, social interaction and mobile communication patterns. The objective is to analyze whether the system is able to recognize significant behavioral changes due to medium-term stress periods.

Many related approaches are based on physiological parameters and skin attached sensors. These sensors may be too obtrusive for long-term application scenarios. By contrast, this approach uses smartphone integrated sensor systems exclusively in order to guarantee an unobtrusive all-day behavior monitoring. The focused research questions were:

- *What kind of behavioral parameters, which are influenced by medium-term stress periods, can be determined by smartphone sensors?*

- *How can behavioral parameters be calculated without compromising the usability of smartphone functionalities?*

The main contribution of this chapter is the proof of concept, that behavioral parameters derived from smartphone sensors are able to reflect behavioral changes due to medium-term stress periods. A multi-user experiment including two weeks of continuous stress and two stress-free weeks showed, that the defined parameters are able to indicate average behavioral changes with a certainty of between 39% and 73%. As already mentioned, behavioral parameters were defined and calculated, keeping in mind that smartphones must still be usable in terms of battery life (the phone should last for more than six hours during normal usage) and provided functionalities such as web browsing, email or making phone calls.

1.4.2 Smart Home Integration

The closing part of this thesis deals with the integration of activity and context recognition services into smart environments. For real-life, large-scale and market-orientated applications the possibility of distributed service developments among different partners, the re-usability of available services, an easy to perform service installation procedure and the possibility to extend exiting systems is of high importance. These points can be easily accomplished if service developers follow standardized interfaces. So the main contribution of this chapter was to define standard components and based on them standardized service interface descriptions for real-life ambient intelligence applications. As it would go beyond the scope of the thesis to consider all application scenarios, this work is restricted to Ambient Assisted Living (AAL) applications in smart home environments[15]. Since the OSGi platform is one of the most used gateway platforms in such scenarios, the defined components and interfaces were designed as an extension of OSGi in order to adapt it to AAL scenarios. The following research questions were observed:

- *Which requirements have to be made for a service platform for smart environments?*
- *Which common components can be defined for AAL services?*
- *How can ambient intelligence systems be integrated into OSGi in a standardized way? And how can the restrictive influence of standardization on the creativity of service developers be reduced?*
- *Can end-user services for real-life scenarios be realized in a distributed way and by re-using existing services when employing standardized interfaces?*

This thesis introduces a three-layer architecture that groups components in low-level sensor devices, technological services and high-level functional (end-user) services. Based on this categorization, a standardization approach (called OSGi4AMI) for common AAL service interfaces is introduced. In this context interfaces were defined for technological services only in order to guarantee a maximum on flexibility for service developers. The resulting set of interfaces includes, amongst others, services introduced in Chapter 2 and Chapter 3. As a next step, it is shown that real-life end-user services can be realized based on OSGi4AMI. Furthermore, the advantage of re-using existing service bundles and distributed service development is demonstrated. Additionally, the end-user acceptance of innovative assistive services is evaluated. Two services related to home security (area surveillance and the monitoring of electronic devices based on techniques shown in Chapter 3) were implemented in a distributed manner and by re-using existing components. Services were installed in real end-user homes and were used by residents (elderly and disabled people) for three months. User questionnaires evaluated by the "London School of Economics and Political Science" (LSE London) showed, that such innovative services were fully accepted by the participants. 71% of the participants stated, that the systems were integrated unobtrusively in their existing homes and therefore assistive services were woven into their everyday lives just like the vision of Mark Weiser in [Wei99]. Besides, 32% stated that these services have influenced their ability to look to the future more optimistically as they know that dangerous situations will either be prevented by the system or that the system will call for help immediately. In summary, the main contribution of this chapter is the fact that OSGi4AMI is a feasible solution for smart home environments to realize distributed service development and the re-usability of existing services. Besides, an initial study showed that the deployed systems based on minimal training data, unobtrusive installations and low-cost sensors are accepted by society and are able to improve people's quality of life. I want to note, that OSGi4AMI has been considered to be integrated in the AAL Open Association (AALOA)[16]. Furthermore, the creation of OSGi4AMI was supported by the EU project MonAMI, EasyLine+ (IST-045515) and by the Spanish Ministry of Science and Technology under the AmbienNET

[15] The decision that AAL applications were considered was mainly influenced by the EU project MonAMI.
[16] http://www.aaloa.org/ (last accessed on 2013/05/10).

project (TIN-2006-15617-C03-02).

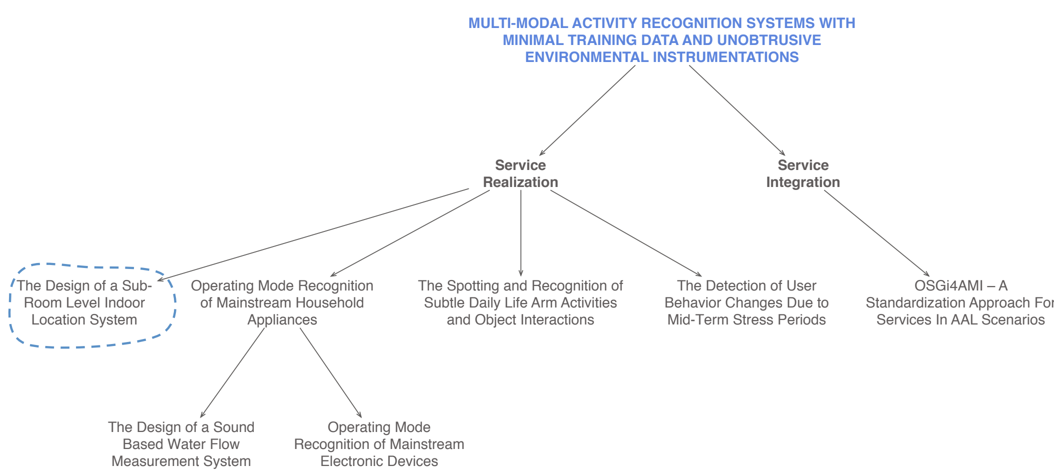
Before going into detail with the main topic of this thesis, I want to note that a complete overview of my publications, which create the basis of the thesis, can be found in the appendix.

2

Design of a Sub-Room Level Indoor Location System

This chapter is based on my work published in

Gerald Bauer and Paul Lukowicz. *Developing a Sub Room Level Indoor Location System for Wide Scale Deployment in Assisted Living Systems*, ICCHP (K. Miesenberger, J. Klaus, W. L. Zagler & A. I. Karshmer, eds.), Lecture Notes in Computer Science, vol. 5105, Springer, 2008, pp. 1057-1064.



2.1 Introduction

Locating people indoors is an issue often addressed in the pervasive computing research community. The location of a person is a core piece of information for many applications and serves as a fundamental component in many sensor fusion approaches. Examples are: Emerging assistive technologies, reminders about essential everyday activities (e.g. medication or hydration) or behavioral assessment systems that aim to detect abnormal behavior patterns from daily routines.

Various location systems have already been successfully deployed within adapted environments and research labs in general. Although some of these systems are able to determine the location of people and objects quite accurately, their use is largely restricted in real-world and large-scale scenarios. The fact that such systems are based on highly instrumented environments and some of them include expensive sensor modalities is clearly the main reason for this limitation. It is obvious that people are neither willing to spend more than 15.000 € on location systems (e.g. the UBISENSE system[17]) nor do they want the intrusion of sensor instrumentations on their interior design.

This chapter introduces a novel approach for localizing people on a sub-room level and works towards ameliorating the problems mentioned. The proposed positioning system is based on minimally invasive instrumentations and uses affordable ceiling cameras (under 500 €) in combination with low-cost wearable motion sensors. High-level motion patterns derived from a standard computer vision based tracking approach and on-body acceleration sensors are correlated. As a result, the approach shown is able to locate people on the basis of pre-defined regions of interests (ROIs). Such regions may include important areas within a home or even the location of items like household appliances, doors, windows or furniture (see Figure 2.1). Thus the achieved location precision is significantly better than room level.

Figure 2.1: Household items, that are interesting for many recognition applications. Living room (left image) and kitchen (right image).

2.2 Related Work

The problem of indoor positioning is complex and difficult (see [HB01]). In general indoor location systems can be grouped into different major categories. The first group includes affordable and easy to deploy systems providing room level location only. The most common approaches are based on fingerprinting in combination with WLAN or other RF technologies as ZigBEE or Bluetooth (see [BP00] [BBC+03] [NDA06] [PCM+06]). In [BP00] a radio-frequency (RF) based system for locating and tracking users inside buildings is introduced. Experiments were carried out within a 980 sqm area including more than 50 rooms. A median resolution of two to three meters was achieved by using three base stations, empirical signal strength measurements as well as simple signal propagation methods. [NDA06] introduces a fingerprinting technique using the channel's impulse response information and artificial neural networks. The system was evaluated in a mine scenario and is able to provide a location accuracy of two meters for 90% of trained and 80% of untrained patterns. Moreover, RFID based concepts were used to provide location at room level (see LANDMARC [JLP06] [NLLP04] or SpotOn [HWB00] [HVBW01]). In [HWB00] the development of SpotON, a fine-grained indoor location system based on RF signal strength is introduced. The system uses low power tags that can be located with an accuracy of $1m^3$. Additionally, tags have enough processing power to provide authentication or caching functionalities. Other approaches utilize PDR systems. Based on acceleration sensors integrated in smartphones or shoes, people can be tracked in a very unobtrusive way (see [Fox05] [BWK08] [KR08] [KLF11]). However, such systems require a detailed floor map and their performance is not as good, when it comes to position accuracy. [KLF11] investigates how

[17] www.ubisense.net (last accessed on 2013/09/15)

an ad hoc collaboration between devices located close to one another, can improve the quality of PDR systems. The system was evaluated with 12 people during a three day open air festival in Malta. It is shown that exploiting collaboration can improve the localization accuracy by a factor of four. Besides, it can prevent unbounded PDR errors. Another quite promising approach is based on magnetic field measurements (see [PL12]). So far this approach is only at a prototype state and requires magnetic coils to be placed in the environment as well as special on-body sensors. These limitations may be too obtrusive for real-life applications. Sound-based approaches have been also used to locate people indoors (see [BAS96] [BAS97]). However, the pre-condition of such location systems is, that people can only be located while they produce noise. As this cannot be assumed in many real-life scenarios, such systems are only usable in specific applications. Vision-based systems have been introduced to locate people on a sub-room level [KHM+00] [SSP98]. Even the issue of person identification using vision-based face detection systems has already been solved. In [HYI04] and [AB96] systems are shown that are able to identify people by using such algorithms. In [AB96] a person identification system is introduced that combines several face classifiers. Beside profile classification techniques, frontal view classifiers as the Eigenface classifier and HMM classifiers have been considered. By combining all of these classifiers, a recognition rate of 99.7% could be achieved. Although such systems are able to provide promising results, they rely on specific conditions that may be not given in many real-life home environments. The most obvious one is that the camera has to clearly see the face of a person.

Many highly accurate positioning systems with an accuracy ranging from a few meters up to only centimeters exist. Examples are ultra wide band systems (e.g. [AAP06] or Ubisense[18]) , ultrasonic systems (e.g. [SBGP04] [ACH+01] [Pri05]), various smart floor approaches based on pressure [OA00] [SFR08] and capacitive sensors [SL08] [STS+13] [GPBB+13] as well as infra-red camera based systems like the Lukotronic system[19]. However, such systems are either excessively expensive (e.g. more than 15.000 € for a one-room UWB system) or the installation and setup procedure is very complex and time-consuming (several nodes have to be placed, extensive calibrations have to be performed or sensor mats have to be placed under the floor). These facts often prohibit a large-scale deployment in existing real-world scenarios.

In contrast to the related approaches, this work introduces an indoor location system (including the identification of people), that offers a sub-room level precision while still being unobtrusive and low-cost. Besides, the system is easy to deploy, to maintain and is even scalable. Consequently, the system is optimally adapted to real-life scenarios, large scale deployments as well as changing environmental conditions. Similar approaches are described in [IIN07] and [TJS10]. In [IIN07] acceleration signals generated by motion sensors carried by hand are mapped to intensity pixel changes generated by simple difference images. The system is based on acceleration sensors built in to Wiimotes. Two initial experiments showed that the proposed system works when identifying two people with swinging arms while walking. However, authors remark that future work has to be done in order to evaluate the impact of more people and various behaviors like "skipping" and "dancing". In contrast to that work, this chapter introduces a system that does not use raw motion data but processed data and resulting high-level activity information. In this way the impact of small motions (e.g. coming from shaking hands) or even wrongly detected motions (e.g. because of illumination changes) can be reduced.

In [TJS10] the idea of my approach is extended. People are also identified and localized by leveraging ceiling cameras along with smartphones with integrated motion sensors. In contrast to this work, the authors additionally consider the walking direction and not just binary motion information (walking / standing). As a consequence, [TJS10] is a clear continuation of my work, also being published two years later.

In [ZS11] motion data coming from an on-body acceleration sensor is combined with several cameras. The objective is to improve the accuracy of an activity recognition system that focuses on the detection of activities of daily life indoors. First, activities are recognized based on motion data only. Afterwards Bayes' theorem is used to fuse motion data and location information.

[18] www.ubisense.net (last accessed on 2013/09/15)
[19] www.lukotronic.com (last accessed on 2013/09/15)

For example, the fact that the subject is on the sofa decreases the probability of "walking" by a lot. The system was evaluated in a mock apartment environment and achieved an overall accuracy of more than 85%. The achieved recognition rate was significantly higher compared to systems using only motion data or human activity recognition methods based on video data. In [BHG+09] a WiFi webcam and standard mobile phones were considered as sensor systems. The visual activity monitoring was improved by electromechanical movement data coming from mobile phones. The system was evaluated in an indoor as well as an outdoor scenario including activities like "car driving", "cycling" or different gestures such as "pointing to the TV".

2.3 Research Questions and Contribution

This chapter focuses on the problem of user identification in camera-based indoor location systems. Computer vision approaches, which are able to detect moving people, to track and to identify them, have already been introduced. However, such approaches are often based on high resolution cameras, are computationally intensive and are very sensitive to changing light conditions or background clutter. In contrast, simple motion and foreground-background detection systems do not have to face these problems. However, in turn they cannot provide user identification. These systems are only able to detect that "something" is moving. Such unidentified objects are called "blobs".

Consequently, the considered ***key problem*** can be described by:

How can a simple, computer vision based motion tracking system be extended to be able to identify tracked persons?

Most obviously, state-of-the art computer vision methods can be used to identify people on images. Widely used approaches are based on face recognition methods as shown in [HYI04] [AB96] [BKYC07]. Such systems were not considered in this work due to the afore mentioned problems in real-world environments (see Section 2.2).

Consequently, the focus of this work was on finding an appropriate sensor modality and fusion methods, which are able to identify tracked "blobs". Besides that, considered models should be based on simple one-time configurations or minimal training data to meet the requirements of large-scale and real-world deployment. This leads to the ***first research question***:

What kind of person-specific features can be derived from unidentified blobs? And which models can be used, that are able to identify the person based on these features and using simple one-time configurations only?

The key information of a vision based tracking system is the position of an unidentified object on an image over a period of time. This work uses a ceiling mounted camera to determine a two dimensional pixel based position of moving objects with one single camera. Section 2.5.2 introduces a method to calculate basic movements (object is moving or not) based on a sequence of the derived position values. Obviously, the same information, but user specific, can be derived from on-body motion sensors. Many approaches have already introduced modes of locomotion detection systems based on on-body motion sensors (e.g. [FMT+99] [JKC07] [RM00] [VLC00] [LM02] [MHS01]). However, due to the fact, that the focus is on binary movement information, a simple, threshold based standing detection algorithm was used in this work. Based on these considerations the ***second research question*** was:

What fusion method can be used to assign user-related on-body sensors to identified objects based on their movements?

Section 2.5.4 defines a correlation method to solve this issue. The key idea is to define characteristic movement profiles for both: tracked blobs and people wearing an on-body motion sensor. The defined profile consists of a sequence including time and type of movement changes. By correlating movement profiles, on-body sensors and hence user identity information was assigned to unidentified blobs. Of course, the main pre-condition of the system is that objects have to show a different movement behavior at least once.

As soon as the moving object is identified, its location has to be determined on a sub-room level. This leads to the ***third research question***:

How can a pixel-based position of an identified blob be mapped to a sub-room level location without much effort?

This issue was solved in Section 2.5.3 by using pre-defined regions of interest (ROI) on images. In fact, a ROI is defined by polygonal lines, which mark interesting regions on an image. Such ROIs can even be defined by technical laymen without much effort. Using standard methods as described in [Kle08], the pixel position of a moving person can be located either inside or outside of such a ROI. Consequently, the current position can be determined on a sub-room level.

The last problem to be considered in this chapter, is related to the usage of standard vision-based tracking systems (used to track motions in general) in indoor, person tracking scenarios and ceiling cameras. This leads to the ***fourth research question***:

What are the problems when using a standard computer vision-based motion tracking system in indoor person tracking scenarios? Which methods can solve these issues with simple one-time configurations?

In general, the following three problems were addressed in Section 2.5.1.1:

- First, movement tracking systems detect every kind of motion, no mater how small. The idea of this work was, to allow only motions caused by moving humans with high probability and to reject other motions such as hand movements. This issue was solved by the definition of an easy to parametrize human model restricting the size of detected blobs.
- Second, the tracking system has problems to assign new, fast movements to detected blobs. Consequently, a new blob is wrongly generated in such scenarios. This problem was solved by a method, which also considers the close circular surrounding of each blob during the assignment process.
- Third, the tracking system deletes blobs (and hence their motion profile), if they have not been updated for a specific time range. Such situations may appear often in real life scenarios (e.g. person sitting on a chair). Due to the issue first described, small person movements cannot be considered during the update process as an indication that the person is still there. This problem was solved by introducing a more intelligent life-cycle approach considering even small motions such as hand movements during the update process.

Finally, the aspect of privacy is a big issue in indoor surveillance applications. This fact leads to the ***fifth research question***:

Can the system proposed be realized in a way, where privacy concerns are weakened?

This issue is solved by introducing a decentralized approach where personalized data remains on user-assigned devices in Section 2.4. Information about unidentified blobs and their motion profiles only have to be transmitted.

In summary, the ***key contributions*** of this chapter and the extension of the current state-of-the-art are:

- Solving the problem of user identification in computer vision tracking applications by correlating modes of locomotion patterns which are derived from blob trajectories with on-body motion sensors.
- The definition of a correlation method based on movement profiles derived from tracked camera objects and on-body motion sensors.

Apart from that, the approach proposed meets the following requirements on real-life, large-scale scenarios:

- The system is unobtrusive, low-cost, easy to install, easy to deploy and highly scalable.
- The system works "out-of-the-box". This means, that only simple one-time measurements and configurations have to be performed to deploy the system instead of using excessive training procedures based on large training data sets.
- The system is able to weaken privacy concerns. This aspect is one of the most important requirements of real-world and large-scale scenarios.

I want to highlight, that although the system shown is based on cameras and computer vision methods, the contribution of this work is **not** in computer vision. Contrarily, this chapter introduces a novel way of combining simple vision algorithms with wearable motion sensors to overcome the problem of user identification. Hence, the key advantage of the system is that adequate location information can be provided without the need for advanced image processing.

I would also like to point out, that the idea of combining computer vision methods with motion pattern recognition systems has been picked up many times over the last few years (e.g. [TJS10] [IIN07] [ZS11] [BHG+09]). However, this work was amongst the first publications in this field and can be seen as pioneering. It has been cited by 13 publications[20].

The following sections describe the system concept, applied algorithms and processing steps and the current implementation using a ceiling mounted fish-eye camera and motion sensors integrated in iPhone/Nokia N95 mobile phones. Finally, initial results achieved while operating the system in real-time and within a lab environment under conditions similar to real-life are reported. It is shown that the system is able to recognize the user's sub-room level location with an overall absolute time delay of 2.1 seconds.

2.4 System Concept

This work uses standard foreground-background detection algorithms to track objects. The detection of moving objects is achieved by simple algorithms which are insensitive to changing light conditions as well as background clutter and are even able to track several moving objects at the same time. Such systems fit perfectly to real-life applications. In those scenarios, multiple users have to be tracked and used systems must be easy to deploy even on a large scale. However, as already mentioned, the price for sensitivity and robustness is the lack of user identification. Even worse, such simple vision algorithms cannot differentiate between human beings and other moving objects (e.g. confusion with pets). Rather, they only output "blobs". This work solves the issue of identification by using on-body motion sensors. Motion patterns for each blob (calculated using their trajectories) are compared with movement patterns coming from on-body accelerometers. In this way, motion sensors and hence user ids are assigned to unidentified moving objects and hence enriching them with identity information. This enables the system to track and identify people in images. Depending on the vision tracking approach, the position of a person can be determined fairly accurately. As this work focuses on the requirements of minimally invasive instrumentations, low-cost sensors and minimal training effort, a rather simple vision approach was chosen.

[20]Source: Harzing's Publish or Perish (last accessed on 2013/05/31)

It requires a single camera to be mounted in the center of the room's ceiling. The camera could be powered by a nearby light outlet and even data transmission could be done using power line techniques. Besides, "Power over Ethernet" could be used to solve both problems with one cable. Consequently, the installation is easy to perform and can be done by electricians. As high resolution images are not needed, even simple and low-cost cameras (like mainstream webcams) can be used. When using a fish-eye camera, the whole room could be monitored quite unobtrusively by one single camera. Hence, the amount of cameras needed is minimized.

The accelerometers do not need to be fixed onto a specific place on the human body in order to determine simple movement patterns. Thus, accelerometer sensors integrated in mainstream consumer products like wrist watches, belts or shoes can be used. This work is based on motion sensors built in to mainstream smartphones, that are simply carried in the user's pocket. Although identification is not an issue for people living alone, it may become important if visitors arrive. In this case, mobile phones of guests could automatically register to the system and hence their generated blobs can be distinguished from the one created by the house owner.

Another important feature of the proposed approach is the fact, that it runs in real time on a low-end PC and could even be implemented on a FPGA (providing a cheap embedded solution).

Concerning the important aspect of privacy, the proposed system provides an easy solution. People can simply decide themselves if and how long they want to be monitored. As soon as they turn off their motion sensors, the system will not be able to identify them (in single user scenarios the system can be configured to insist on the existence of motion sensors). Besides, in theory, the system does not need to store or transmit video frames. Only information about unidentified blobs, their current mode of locomotion and their sub-room location need to be sent to a central processing unit. There, data coming from motion sensors (current motion pattern and identification) are compared. If cameras and motion sensors can be connected, the comparison can even be done on the user's mobile, providing greatly increased privacy. Figure 2.2 shows two different architectures realizing the system introduced, whereas one solution (Figure 2.2 (b)) is able to overcome many privacy concerns.

(a) Central processing unit.

(b) Image processing on FPGA and blob assignment on smartphones.

Figure 2.2: System architecture. Figure (a) shows a centralized solution based on a single central processing unit. In Figure (b) a decentralized solution is shown using FPGAs and smartphones as computation units.

2.5 Algorithms and Methods

This section explains the applied processing tasks of the system proposed in detail. An overview of the computation tasks performed is shown in Figure 2.3. In the following included components are discussed step by step.

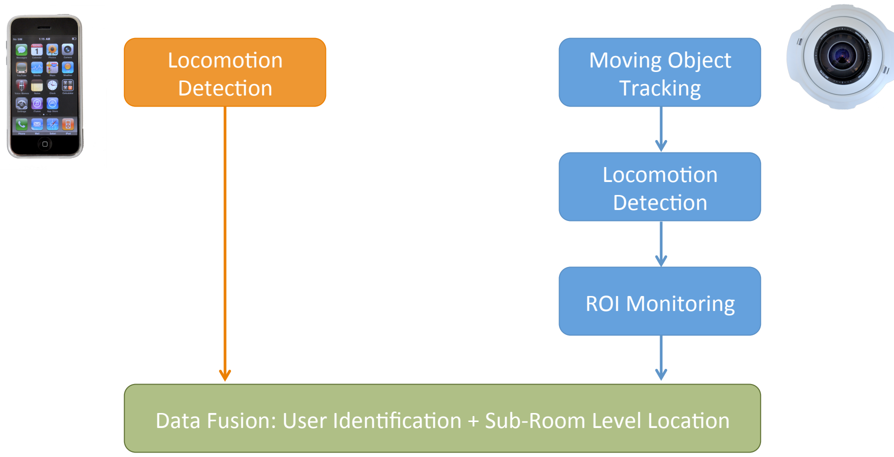


Figure 2.3: Data processing chain. Processing steps for smartphone data (left chain) as well as images coming from surveillance cameras (right chain) resulting in a final fusion step.

2.5.1 Moving Object Tracking

The detection of moving objects is based on a standard blob tracker module (which is included in OpenCV[21] and from now on is called SM - Surveillance Module). Given an image stream, SM detects moving objects (so called "blobs") as well as their trajectories. Each blob is described by the object's center and its size (height and width). This information is used to calculate the current mode of locomotion and the sub-room level location of detected blobs in the following steps. Figure 2.4 shows an example of a detected blob (green circle) and a blob's trajectory. A list of existing blobs (with an internal, unique blob id ($blob_{id}$)) as well as information about their size and current position (pixel-based) are provided to the next computation layer. The system is able to process images at a sampling rate of about 8 Hz.

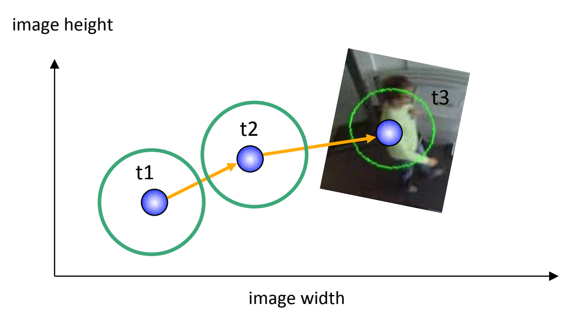


Figure 2.4: Left: Blob with blob identification number 000 detected (green circle). Right: Blobs are tracked on images t1-t3.

During initial experiments, it turned out that the SM in its original version is not able to reliably solve the task of detecting and tracking people indoors using ceiling cameras. As a consequence, it was adapted to handle issues related to small hand movements, loosing blobs or wrong blob allocations. A detailed problem description and related solutions are shown in the following.

[21] http://opencv.org/ (last accessed on 2013/08/28)

2.5.1.1 Blob Tracker Module and Problems Related to Indoor Human Tracking Scenarios

The SM provides a general solution for moving object detection and tracking. Since indoor scenarios, the usage of ceiling cameras as well as the exclusive tracking of humans exact special requirements from tracking systems, the original SM was extended with an additional layer called $blobHandler$ to deal with the following problems:

- **Small Movements:** The SM detects every kind of movement, however small. Consequently, even very small movements such as arm motions are recognized as stand-alone moving objects. Image 2 on Figure 2.5 shows an example. As the focus of this work is to detect only moving people, the $blobHandler$ filters out blobs based on small movements. In detail, each blob having a height and width of less than n pixels was rejected. In the scenarios observed and at an image resolution of 320x240 pixels, $n = 12$ worked very well for several different experiments. Hence, the following evaluations are based on this configuration.

- **Wrong Blob Allotment:** Initial experiments showed that if a person suddenly starts walking fast after standing for several seconds, the SM is not able to assign the existing blob to the new movement. So, a new blob is wrongly generated while the old blob is still alive. Image 1 on Figure 2.5 shows exactly the moment where a fast movement results in a tracking failure. As this is a common situation in indoor scenarios (e.g. people leave the room after they have been sitting for a long time on couches or chairs), the $blobHandler$ merges new unassigned blobs to existing ones, provided that they are located within a range of m pixels. Initial experiments have shown that a value of $m = 80$ works well for the considered scenario.

- **Lost Blobs:** The SM assigns blobs to the background if they do not move for a certain time (specified by a threshold). Once a blob is assigned to the background, all its information as well as the corresponding trajectory are deleted. In indoor scenarios (e.g. person watching TV) such situations are especially quite common. To overcome this problem, the threshold can be set to a very high value. However, this will also lead to a prolonged retention of falsely detected blobs. Consequently, a more adjustable lifetime approach was chosen. A blob and its information is only deleted if it has not been updated during the last few frames. However, disregarded small movements (like hand motions) are also used to update the life-cycle. This is done by a counter for every blob. Initial experiments showed that a starting value of 10, an increase of 1 (if movement could be assigned to an existing blob) and a decrease of 5 (if no movement could be matched to an existing blob) worked well for this scenario.

A remaining unsolved problem are people located close to each other. In some cases, blobs cannot be assigned to the correct people and consequently they are mixed up. Image 3 and Image 4 on Figure 2.5 show two examples. Trajectory histories from two different people may be confounded. There are many approaches dealing with problems like this or reducing their negative impacts. However, this work's focus is not on developing the best camera-based indoor multi-user tracking system. Rather it aims to show, that on-body motion sensors can be used to enrich mainstream vision-based tracking and sub-room location approaches with user identification information. For this reason no further attempts to improve the vision-based computation chain were undertaken.

2.5.2 Locomotion Detection

As camera and motion-based systems provide different types of data, different approaches aiming at the recognition of so-called modes of locomotion must be considered. Although modes of locomotion in general describe many different types of motion (e.g. "sitting", "walking upstairs" or "running"), this approach only uses two of the most basic ones, "walking" and "standing".

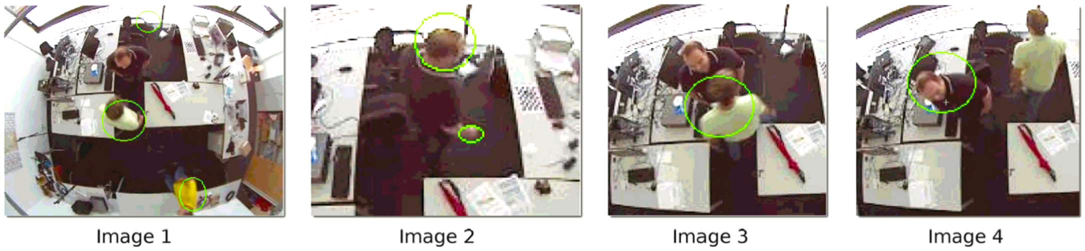


Figure 2.5: Object tracking module and problems with indoor and human tracking scenarios. Image 1 - Image 2 show wrong blob detections and Image 3 - Image 4 wrong blob assignments.

First, the camera-based system is focused on. By analyzing the trajectory of a blob over short periods of time (as already shown on Figure 2.4 (right image)), it can be determined whether the person is currently walking or standing. In detail: For each frame, the Euclidean distance between the center of a blob on the current and the last frame was calculated. A sliding window of size eight is applied to the time series of these values in order to calculate variance values. For values larger than 17.86 (this threshold was chosen based on initial experiments) the person is assumed to be "walking" and otherwise "standing". In order to reduce the impact of false classifications and outliers, a majority decision is performed on the last five calculated modes of locomotion. The result is chosen as the current activity.

The recognition of modes of locomotion based on acceleration sensors is a well-researched topic (e.g. see [RM00] [LM02] [MHS01]). This work uses a standard procedure to recognize binary motion activities ("walking" / "standing"). In the following, the algorithm is explained for the Nokia N95. First, variance values were calculated using a sliding window of size 10 on the Euclidean norm of 3-dimensional acceleration vectors[22]. In order to detect modes of locomotion, the last 100 variance values were summarized. Initial experiments showed that a value below 10 is a very good indicator for walking activities. Consequently, the person is assumed to be "standing" for values below and equal to 10. A smoothing function was used to reject short but intense motions. A walking activity is assumed to be if 90 percent of the last 100 classified instances have been of type "walking". The mode of locomotion detection for iPhone devices is done in a comparable way. Both approaches are based on the assumption that mobile phones are carried in trouser pockets.

2.5.3 ROI Monitoring

Simple region of interest (ROI) surveillance approaches based on motion detection techniques are already included in many mainstream camera systems (e.g. Axis 212PTZ). However, very often such systems are not able to assign motions to specific objects. The objective of this work was to locate several people on a sub-room level. Those regions themselves have to be defined by the user once. Examples could be areas in which household appliances (e.g. a coffee machine), doors, windows or furniture are located. Figure 2.6 shows pre-defined regions of interests on a camera image. There, a shelf, the coffee area as well as the printer device were marked as important points of interest. The ROI monitoring task checks if the center of the available blobs lies within the pre-defined ROI polygons using standard algorithms as described in [Kle08]. The regions of interest discovered and hence sub-room level location information is assigned to each blob.

2.5.4 Data Fusion: User Identification and Sub-Room Level Location

So far there are still two separate processing chains. The first, an acceleration-based one delivers information about the current mode of locomotion of known persons (via their individual

[22] Acceleration data was processed with a sampling rate of 38 Hz.

Figure 2.6: Camera image with three ROIs (green polygons – shelf, coffee and printer area).

sensors). The second, a video-based processing chain provides unidentified moving objects, their current mode of locomotion as well as sub-room level location information. Now, the fusion step combines both sources of information. Each unknown blob is assigned to a mobile phone and thus identified. The assignment is based on the correlation of two individual modes of locomotion datasets. For smartphones and camera blobs, the last k timestamps (t) when the type of motion changed and the resulting new motion type (act) were stored. The following distance measure was used to assign blobs ($blob$) to motion sensor devices ($accDev$):

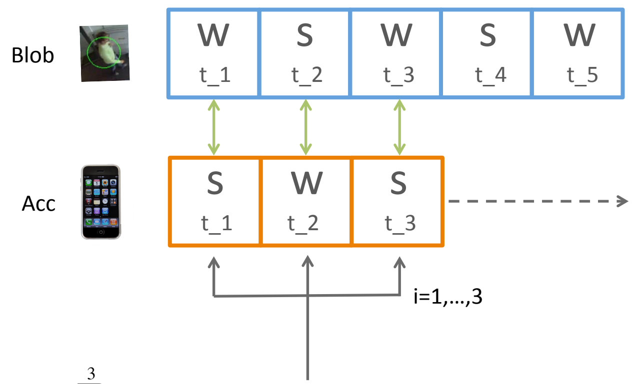
$$dist(blob, accDev) = max(distActSeq(0), ..., distActSeq(a - b)) \tag{2.1}$$

$$distActSeq(n) = \begin{cases} \sum_{i=o}^{b-1} comp(accDev_{act_{i+n}}, blob_{act_i}), & \text{if } (||accDev_{act}|| > ||blob_{act}||) \\ \sum_{i=o}^{b-1} comp(accDev_{act_i}, blob_{act_{i+n}}), & \text{otherwise} \end{cases} \tag{2.2}$$

$$comp(act_i, act_j) = \begin{cases} 1, & \text{if } ((act_i == act_j) \wedge (abs(t_i - t_j) < thr_{time})) \\ 0, & \text{otherwise} \end{cases} \tag{2.3}$$

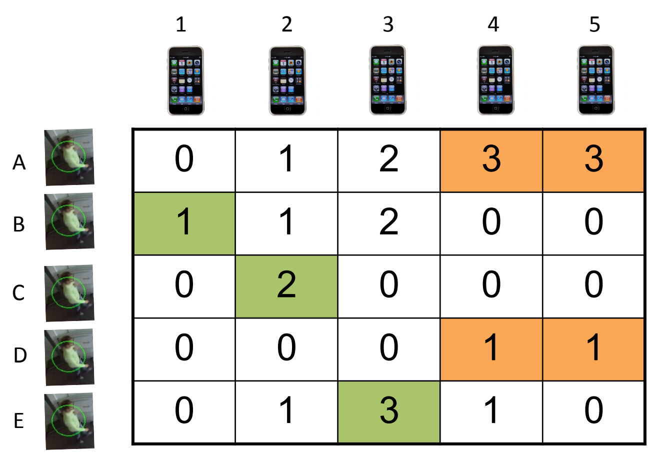
There, a and b are the maximum and minimum number of stored activities in terms of a specific blob-motion sensor pair. Consequently, $0 \leq b \leq a \leq k$. The introduced functions have the following domains: $dist : (blob \times accDev) \rightarrow \mathrm{N}$; $distActSeq : \mathrm{N} \rightarrow \mathrm{N}$; $comp : (act \times act) \rightarrow \{0, 1\}$. Initial experiments have shown, that a thr_{time} of 2.5 seconds is sensitive and robust enough in terms of timing and synchronization. Figure 2.7 shows an example in order to visualize the idea of $comp(act_i, act_j)$. A blob with five registered locomotion changes is compared to a mobile phone locomotion history list of size three. For each hit, which means both having the same locomotion type as well as a similar modification time, the current assignment is valued with one credit. If the end of one considered pattern is reached, the smaller pattern moves one step to the right and the comparison restarts. The algorithm finishes if the end of the smaller pattern reaches the end of the larger one. The highest achieved value is used as a correlation value of the pair under consideration. Finally, all combinations between smartphones and available blobs are compared and pairs with the highest rated values are assumed to belong together.

Figure 2.8 shows the assignment procedure based on correlation values already calculated for five smartphones and five blobs. In this case, three smartphones (1, 2 and 3) could be assigned to blobs. The remaining two devices could not be matched, as the corresponding correlation values were not unique. This implies, that these two people carried out similar movements at a similar time (e.g. walking in tandem, stopping to chat, etc.). But even though the system is not able to do an assignment at this time, as soon as one person performs a different motion ("standing" / "walking"), the system will even be able to identify these blobs.



$$\sum_{i=0}^{3}(if\ ((act_{i_{BLOB}} == act_{i_{ACC}}) \wedge (abs(t_{i_{BLOB}} - t_{i_{ACC}}) \leq 2.5\,\mathrm{sec}))\ then\ 1, else\ 0)$$

Figure 2.7: Correlation calculation between a blob's and a motion sensor's locomotion trajectory (s=standing, w=walking).



	1	2	3	4	5
A	0	1	2	3	3
B	1	1	2	0	0
C	0	2	0	0	0
D	0	0	0	1	1
E	0	1	3	1	0

Figure 2.8: Example of blob-motion sensor assignments: Correlation values for each mobile phone – blob pair. Mobile phones four and five cannot be assigned to blobs at this point as the correlation values are not unique (orange colored). Smartphone one was assigned to blob B, phone two to blob C and phone three to blob E.

2.6 System Implementation

This section describes how the proposed system was realized. The overall system architecture follows the concept described in Figure 2.2 (a). A central processing unit is used to analyze images, receive motion patterns from on-body sensors (via WLAN) and to perform the correlation. In the following section, hardware and software tools, which have been considered, are explained.

2.6.1 Hardware

The system is based on two sensor modalities only: ceiling cameras and on-body motion sensors.

2.6.1.1 Ceiling camera

Almost any mainstream camera fits the requirements of the described system. The implementation shown uses a wide-angle camera from Axis (Axis 212 PTZ) that was mounted in the center of the ceiling. Thus the system is able to cover a whole standard-sized room (up to about 36 sqm in the case of common ceiling heights) with a single camera only. The camera was powered by PoE (Power over Ethernet), which means, that the power supply as well as image transmission were done using a single cable. This minimized the amount of sensors, environmental instrumentations, cost and deployment effort. The camera used is able to deliver up to 30 frames per second at a resolution of 640x480 pixels. As high-resolution images are not required, the camera was configured to capture images with a resolution of 320x240 pixels. These images were sent to a central computation unit via Ethernet as shown in Figure 2.2 (a).

2.6.1.2 On-Body Motion Sensor

In this work, motion sensors integrated into mainstream smartphones have been used. Todays mobiles are already widely used and known as very unobtrusive sensor systems. For this system, the iPhone (1st generation) and the Nokia N95 were considered, as these devices were the most popular smartphones at that time. Both devices are able to provide three dimensional acceleration values (e.g. the Nokia N95 acceleration sensor is able to provide motion data[23] with a sampling rate of about 38 Hz) and are powerful enough to process raw sensor data. After evaluating high-level motion data, timestamps as well as unique device ids were sent to the central processing unit via WLAN.

2.6.2 Software

The system was realized based on two open source software tools: The CRN Toolbox and the open computer vision library OpenCV. Both tools are described in the following.

2.6.2.1 CRN Toolbox

The "Context Recognition Network (CRN) Toolbox" (see [BAL08]) framework was used to model the processing chain of the system. As Chapter 3 gives a more detailed introduction to the framework's main advantages, only a brief survey is given here. The CRN Toolbox (CRNT) is a software tool which was designed for POSIX operating systems and is optimized for the implementation of multi-modal and distributed context recognition systems. Due to its multi-layer architecture, a modular system implementation and hence a clear separation between different processing tasks could be realized. Computation chains related to raw motion data and image processing are performed independently and only fused at the end. On top of this, due to the fact that the CRN Toolbox can be operated on various POSIX based systems, it is guaranteed that the system can be ported easily (e.g. to smartphones).

2.6.2.2 OpenCV

As the system relies on standard foreground-background detection and tracking algorithms, it re-uses components from the open computer vision library OpenCV (see [CHB+05]). The separation between foreground and static background and hence the detection of moving objects is based on the standard OpenCV "Video Surveillance / Blob Tracker Facility" module. The SM was integrated into the CRN Toolbox processing chain and was combined with external components (not included in OpenCV; explained in the previous sections) to process images, to detect modes of locomotion and to realize a real-time, sub-room level monitoring system.

[23] A value of about 300 corresponds to 1g

2.7 Experimental Results

The proposed sub-room level location system was evaluated by the following experiment, which was carried out within a research lab under conditions close to real-life:

A single fish-eye camera was mounted in a standard-sized public lounge. Besides common furniture (e.g. tables and chairs), the room includes a coffee machine and a laser printer. The area around the coffee machine was marked as a region of interest. Two people were asked to enter the room one at a time. One person (person 1, wearing a green shirt) was instructed to make a cup of coffee, whereas the other one (person 2, wearing a red shirt) was asked to fetch printouts from the laser printer. Both people carried a Nokia N95 mobile in their pocket. The system was used to answer the following question: *Who entered the area around the coffee machine?*

The timeline on Figure 2.9 visualizes the activities performed and Figure 2.10 shows the predefined region of interest as well as some camera snapshots (at the beginning of the experiment, after two and after six seconds).

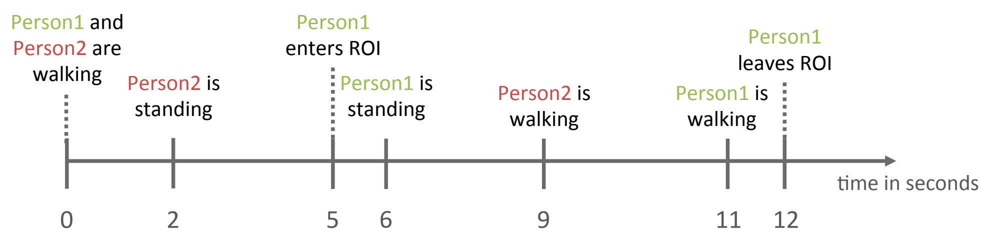


Figure 2.9: Timeline in seconds (after the experiment start): Performed ground truth activities.

Figure 2.10: Sub-room level location evaluation: Blob Tracking. Image 1 shows the monitored region of interest around the coffee area (red polygon) and Image 2 to Image 4 camera snapshots. Recognized blobs are marked with green circles.

The system was used exactly as described in the former sections. Locomotion patterns were calculated on the basis of smartphones and detected blob trajectories. Figure 2.11 visualizes recognized locomotion changes as well as the achieved sub-room level monitoring results. Both persons changed their locomotion mode twice (from "walking" to "standing" and from "standing" to "walking"). Table 2.1 shows the recognized changes and related time values. Although the system was able to recognize locomotion changes, it also exhibits a not insignificant time delay between 1.3 and 3.7 seconds (on average 2.2 seconds). The time delay between the recognized mode of locomotion changes of corresponding pairs, however, is much smaller (0.3 seconds on average for person1 and blob1; 0.65 seconds on average for person2 and blob2). This is a very important fact for guaranteeing a successful correlation process. Each time a new mode

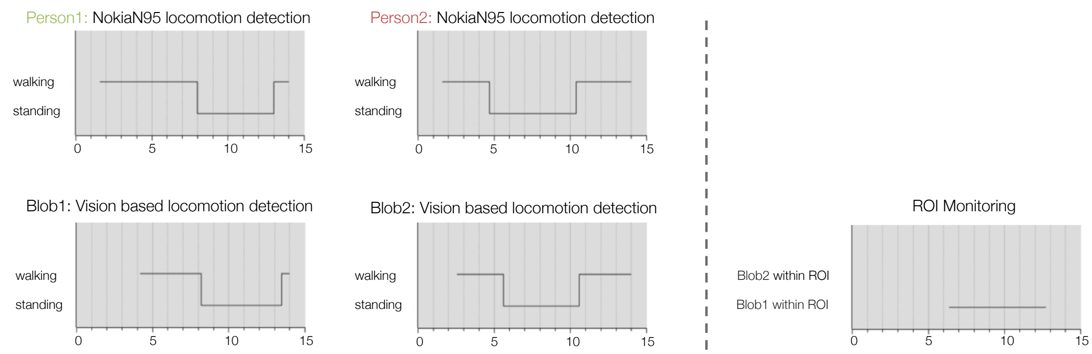


Figure 2.11: These charts visualize the system's output related to mode of locomotion changes and the sub-room level monitoring system. Horizontal axis labels show the elapsed time in seconds.

Table 2.1: Recognized mode of locomotion modification times (in seconds after the start of the experiment) for detected blobs and mobile phones compared to ground truth (in brackets).

	walking $\rightarrow$ *standing*	*standing* $\rightarrow$ *walking*
Person1	8 (6)	13 (11)
Person2	4.7 (2)	10.3 (9)
Blob1	8.2 (6)	13.4 (11)
Blob2	5.7 (2)	10.6 (9)

of locomotion change was recognized, the system tried to correlate unidentified blobs to mobile phones. Table 2.2 shows correlation scores determined by the previously presented algorithm.

Table 2.2: Correlation scores: Calculated for the first four locomotion modifications ("-" indicates that there is no registered mode of locomotion change).

Person / Blob	4.7sec	5.7sec	8sec	8.2sec
1 / 1	-	-	-	1
1 / 2	-	-	1	1
2 / 1	-	-	-	0
2 / 2	-	1	1	1

After 4.7 seconds the first locomotion change is recognized. The system detects that Person2 changed his mode of locomotion from "walking" to "standing". However, a correlation is not possible at that time as there is only one locomotion change so far. After 5.7 seconds, the system recognizes that Blob2 has changed its motion type from "walking" to "standing". Consequently, patterns derived from Blob2 and Person2 are compared. As both objects show the same locomotion change and the time deviation of one second is below the pre-defined threshold, the correlation is scored with one credit. Since more correlations cannot be made so far, Blob2 is identified as Person2. After 8 seconds, the system recognizes that Person1 has changed his/her locomotion mode from "walking" to "standing" as well. When comparing Person1 with Blob2 it can be seen that the proposed algorithm also scores this combination with one credit (the same type of locomotion change and a small enough time deviation between the two motion changes). However Blob2 is already identified and hence no assignment can be made at this time. After 8.2 seconds the system recognizes that Blob1 has changed its motion type from "walking" to "standing". Again correlation values are calculated. At this point Blob1 can be identified as

Person1 as the correlation between the two is scored with one credit and due to the fact that Blob2 has already been identified as Person2.

Figure 2.11 shows, that the system was able to recognize that Blob1 has entered the region of interest after 6.4 seconds and stayed there for 6.3 seconds. We saw that the system identified Blob1 as Person1 after 8.2 seconds. Thus, the system is able to recognize that Person1 entered the region of interest and hence the system was able to locate Person1 on a sub-room level almost in real time. In this scenario, the system was able to locate the person with an overall absolute time error of only 2.1 seconds.

2.8 Conclusion

This chapter discussed a novel way to locate people at a sub room level. The system uses standard computer vision approaches, which are able to quite reliably and efficiently detect moving objects in pre-defined ROIs. However, such systems are not able to identify the person. This work overcomes the problem by sensor fusion with on-body motion sensors. Sensors were used to detect modes of locomotion changes and to correlate them with motion patterns derived from object trajectories.

The system shown fulfills requirements for large-scale and real-world scenarios. This includes the possibility of integrating the system into existing environments in a very unobtrusive way (it only requires cameras to be mounted on room ceilings and mobile phones to be carried by users). Additionally, the system does not need any extensive training data sets. Contrary, only some simple one-time configurations have to be performed (e.g. user – mobile assignment or simple threshold adaptions). Besides, the system is highly scalable; new smartphones only have to be registered once.

An initial experiment showed that the proposed system works quite well with two people in a close to real-world scenario. There, the approach shown was able to locate a person on a sub-room level with an overall absolute time error of 2.1 seconds.

It is clear, that the system faces more problems in crowded places. On the one hand, the simple vision-based approach has to be adapted to handle the problem of many people in close proximity. On the other hand, more locomotion types than "standing" and "walking" must be considered to make evaluated motion patterns more distinguishable. Furthermore, the movement direction can be taken into account in order to get more detailed motion information. This idea was presented two years later in [TJS10].

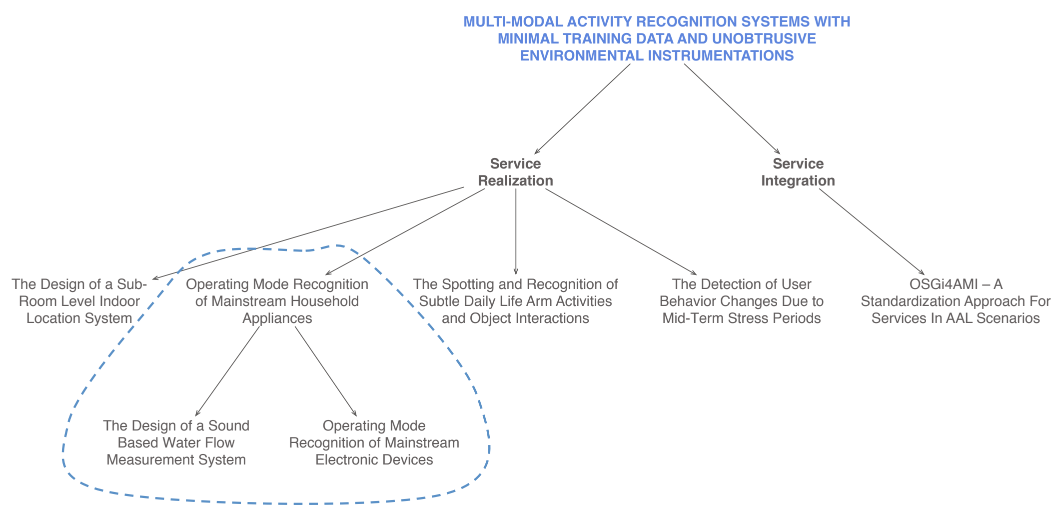
3

Operating Mode Recognition of Mainstream Household Appliances

This chapter is based on my work published in

Gerald Bauer, Karl Stockinger and Paul Lukowicz. *Recognizing the Use-Mode of Kitchen Appliances from their Current Consumption*, EuroSSC (Payam M. Barnaghi, Klaus Moessner, Mirko Presser, and Stefan Meissner, eds.), Lecture Notes in Computer Science, vol. 5741, Springer, 2009, pp. 163-176. – ***Best Paper Award***

Alejandro Ibarz, Gerald Bauer, Roberto Casas, Alvaro Marco, and Paul Lukowicz. *Design and Evaluation of a Sound Based Water Flow Measurement System*, EuroSSC (Daniel Roggen, Clemens Lombriser, Gerhard Tröester, Gerd Kurtuem, and Paul J.M. Havinga, eds.), Lecture Notes in Computer Science, vol. 5279, Springer, 2008, pp. 41-54. – ***Best Commercial Potential Award***



3.1 Introduction

Human activity recognition as well as the recognition of every-day household activities were and are still important issues in many pervasive computing applications. One of several considered scenarios are ambient assisted living applications for elderly and handicapped people. Such systems aim to support people in a way that they can handle their daily life independently or with less human support. Additionally, such systems should prevent people from moving prematurely to overfilled and expensive nursing homes.

The information about which electronic household devices were used and what they were used for is very valuable for such application fields. Besides preventing people from dangerous situations that emerge by leaving on ovens or irons, information about how devices were used are the basis for many complex behavior recognition and abnormal behavior detection systems. Examples are systems designed for the automatic detection and monitoring of a person's course of disease (e.g. dementia) or the detection of abnormal device usage. The manifold possibilities of such systems show their importance for real-life applications and their benefit for problems we have to face today.

This section focuses on two different approaches, that are used to monitor the operating modes of common household appliances. First, a system which is able to approximate the amount of water flowing through a particular pipe (e.g. a water tap or a shower pipe) is shown. The system is based on a low-cost, low quality off-the-shelf microphone module. The microphone was attached to inflow pipes of common water taps and water flow sound samples were analyzed to approximate the amount of water consumed. Second, a new sensor (*iSensor*) was designed to calculate the electric current of the connected devices. The operating modes of electric devices and how they were used was recognized based on rule sets and power profiles.

3.2 Related Work

As monitoring the use of objects is a quite important topic in pervasive computing applications, there is a large amount of research work in this field. The so-called infrastructure mediated sensing (see [PRA08]) has gained in importance enormously as an unobtrusive, easy to deploy and to maintain method for activity recognition scenarios. Examples are the HAVAC system (a motion tracking system based on pressure changes; see [PRA08]), various approaches aiming at water monitoring and electric current sensing systems. The latter two application areas are considered in this chapter.

3.2.1 Monitoring of Water Taps

When it comes to water monitoring applications in home scenarios, many reliable and mid-price sensors are already available on the market. Such sensors measure the current water flow with a very high accuracy and also deliver information about the accumulated water consumption. However, such mainstream systems have two serious disadvantages: First, sensors have to be mounted between the pipes. This fact implies that the water supply has to be broken during the installation process and hence many people, especially elderly and disabled people, are not able to install such systems themselves. Hence plumbers have to be hired and depending on the amount of water taps present in standard homes, the installation procedure can be very expensive and time consuming. Apart from this, so far almost all sensors show the amount of water used on an integrated display and only a few are able to relay this information to a remote device. This fact makes them unusable for smart home applications. The reason is that open systems and the possibility to share information are of great importance for context recognition or home monitoring services. Consequently a lot of research was done aiming at the development of central or distributed water measurement systems. In [CZKS05] an acoustic bathroom monitoring system is introduced. Based on sound analysis, different activities of daily life within a bathroom scenario and related to water usage are recognized. In contrast to this work, authors are recognizing high level activities and are not able to measure the amount of water consumed. Moreover, the system uses a wired microphone, which implies less scalability and higher installation effort. In [KMSW05] various feature extraction methods for sound classification are introduced. Based on the recorded high quality audio in a kitchen environment, it aims at recognizing acoustic kitchen events including water usage, boiling or microwave beeps. Beside the ability to measure the amount of water, the system shown in this work uses only low-cost microphones (mainstream Bluetooth headset). A prototype of a wearable user activity recognition system, which is also based on sound analysis is introduced in [SLT04]. The authors focus on low-power and analyze tradeoffs between recognition performance and computation complexity. The system achieved good results within a kitchen environment and for activities like using the water tap, microwave or recognizing the hot water nozzle. Finally, [FAH06] shows promising work based on a low-cost, wireless and unobtrusive sensor system for sound-based monitoring of water usage. The system is used to recognize many water related activities like using a dishwasher, toilet or washing machine. In [TSGM10] the detection of water flows in hand-washing tasks were considered. There, audio features as Spectral Centroid, Spectral Flux and Mel-Frequency Cepstrum Coefficients were combined with computer vision techniques. The setup consists of a camera placed over the sink. The objective was to recognize water flow variations based on sink detection, hand tracking and temporal features. The system was evaluated using 16 videos (with an average length of 2 minutes) recorded over a four month period. Various classifiers such as kNN or SVM and different feature sets were considered. The best result was obtained by combining all features (audio + video) and using logistic regression for classification. In this way an average accuracy of 88.76% was achieved. [VSN+11] shows a sound based approach which aims at detecting water waste from water tap usage. Two types of water waste were considered. The first one is called inter-activity waste and occurs from free water flows when users forgot to turn off the water tap. In contrast to that, intra-activity waste is defined as wasteful water that does not relate to an ongoing sub-activity (e.g. running water while applying soap). Considered sub-activities were face-washing, tooth-brushing, hand-washing and dish-washing. First, the system distinguishes between running water and a closed water tap. At the next level the system distinguishes between inter-activity waste and user activities. In the case of a user activity the related sub-activity or the intra-activity waste is recognized. The system uses an off-the-shelf headset which was attached to the faucet and the water level was determined by the faucet's opening angle. The experiment consists of three repetitions of each activity. All in all a classification rate of 100% could be reached for inter-activities and 81.1% for intra-activities.

3.2.2 Monitoring of Electronic Devices

Beside water monitoring, electric current sensing is a well elaborated research field. Nowadays only a few mainstream electronic devices are able to provide information about when they were used, how they were used as well as general state information. Examples are intelligent kitchen appliances like smart ovens (e.g. Miele@Home[24]) or smart office devices (such as the Ricoh Aficio SP 5210SR printer). So far, smart functionalities are only included in high-end products that use closed and proprietary protocols. In order to bring assistive services into everyday households on a large scale, sensor systems must be based on mainstream appliances and use open interfaces. Consequently, an affordable and subsequent way to turn common electronic devices into smart devices was one of the main objectives of this chapter. In [BGGBN08] a system is introduced that uses a single sensor to measure electrical events throughout the home. Authors aim at the recognition of high-level daily life activities such as "feeding" or "grooming" by defining the relationships between electronic device usage events, location and time. In [PRK+07] a single sensor is used to analyze transient noise in order to identify which device was used. In [LHP+07] electrical current flow sensors were installed on 37 residential circuits which were used among other sensor modalities in a highly instrumented environment to detect activities such as "listening to the radio" or "using a computer". The easiest way to turn mainstream electronic devices into smart devices is to use power measurement sensors. In [LFO+07] single sensors were connected between power sockets and electronic devices to

[24] http://www.miele-at-home.de/de/aktion/mieleathome/656.htm (last accessed on 2013/09/15)

recognize how the sensor was used. This work uses a similar technique. A power measurement sensor (*iSensor*) providing electronic parameters of connected devices was designed to detect not only *which* device was used, but also to analyze *how* the device was used. Consequently, this work is a clear continuation of related approaches. The issue of energy prediction based on the resident's activities was covered in [CDC10]. There, the hypothesis that energy usage can be predicted based on sensor data is validated. An apartment including a bathroom, a kitchen and a living room was equipped with several sensors (e.g. motion, temperature and water). The environment was monitored over several months. There, all in all six activities were considered (e.g. "work at computer", "sleep" and "cook"). Features such as the length of activities and classifier paradigms such as SVM or multi-layer perceptrons were used to predict the energy use. The highest achieved accuracy was around 90% for a two-class and around 60% for a six-class energy usage.

3.3 Research Questions and Contribution

The recognition of what device type was used within real-world environments was already been addressed by many research projects. Besides the recognition of the common use of electronic devices (e.g. [LHP+07] [LFO+07] [PRK+07]), even water taps (e.g. [FAH06] [SLT04] [KMSW05]) have been focused on. Both, electronic devices and water taps are frequently used daily-life devices and hence, they can contribute to analyze user behavior patterns or to recognize complex activity recognition tasks related to everyday-life events such as cooking or household cleaning. However, approaches aim to detect only ***which*** kind of device was used. This work goes a step further and aims to detect even ***how*** devices were used.

The ***key problem*** which is addressed in this chapter is:

How can common household appliances be subsequently turned into smart devices, which are able to provide their current operating mode (on top of the fact that they are used somehow), with low training and deployment effort?

This chapter introduces solutions for both, electronic devices as well as water taps. The idea of monitoring water taps and electronic devices has also been investigated by other authors over the last few years. However, this work was amongst the first publication in this field and can be seen as a kind of pioneer work. The approaches shown were cited all in all by 31 publications[25] during the last few years and were awarded ("Best commercial potential award" and "Best paper award") as well.

In the following, related research questions and contributions are shown separately for each topic.

3.3.1 Operating Mode Detection of Common Water Taps: Approximation of Water Consumption

In general, electronic devices provide several use-modes. In contrast, possible operating modes of water taps are quite restricted and contain only three kinds of basic information: "Tap turned on", "Tap turned off" and the "Amount of water consumed". Several approaches have already solved the problem of recognizing flowing water in general by identifying water sounds (see [CZKS05] [FAH06] [SLT04] [KMSW05]). This fact leads to a more detailed problem description:

Can the amount of water consumed be approximated by sound analysis as well?

Obviously, the amount of water used can be approximated by summarizing the amount of water consumed between a "tap turned on" and a "tap turned off" event and by considering the current water flow level. This leads to the ***first research question***:

[25] Source: Harzing's Publish or Perish (last accessed on 2013/05/31)

Which features and methods can be used to recognize water flow levels based on sound analysis and minimal training effort?

Section 3.4.2 defines a chain of mainstream audio processing methods and machine learning tools, which are able to solve this problem in real-time by analyzing water samples recorded by a low-cost microphone attached to a water pipe. Three classifier paradigms (with different complexities) were evaluated for their ability to distinguish between seven different water flows based on a minimal set of reference flow samples. The achieved recognition rates on a separate test set showed, that the proposed systems are able to achieve classification rates between 65% and nearly 91% (depending on the classifier used).

This fact leads to the ***second research question***:

Which methods can be used to calculate the amount of water consumed by using only discrete water flow level information? How can surrounding noise be detected and filtered out without reference samples?

Section 3.4.2.3 will introduce an approach using a simple classifier paradigm in combination with a threshold-based method, which is able to detect and to filter out surrounding noises without the need of large training data sets. Besides that, an additional rule layer was defined on top of the classification procedure to approximate the amount of water used. The proposed rule layer includes the following components:

- Outlier detection method
- Algorithms to reduce the impact of false classifications on the approximation result
- Methods to handle predominate surrounding noise

Finally, the ***third*** and last ***research question*** is:

How well can the amount of water consumed be approximated in real-life environments?

Section 3.4.4 shows system evaluations in various real-life environments that were performed by real end-users. Besides the impact of water running through the pipes within the close vicinity of the monitored pipe, the effect of surrounding noise was considered. Finally, a real cooking process was evaluated.

In summary, the ***key contributions*** of this chapter were:

- An unobtrusive, audio based system, that is able to detect running water and to distinguish between various water flows. The system is based on minimal training effort and can be operated in real time.
- The definition of methods (rule layer) that are able to approximate the amount of consumed water, to detect outliers, to reduce the impact of false classifications and to handle drowning noise.
- An in-depth evaluation with real end-users within real environments.

3.3.2 Operating Mode Recognition of Mainstream Electronic Devices

This work identifies devices in a similar way as shown in [PRK+07]. The key idea is to recognize the current operating mode based on an identified device and power features related to the current use-mode. An unobtrusive, low-cost sensor system providing power consumption information was not on the market at the time this work was done. Consequently, an unobtrusive and wireless sensor (the *iSensor*), that is able to calculate the electric current of connected devices, was developed in cooperation with Karl Stockinger (ESL, University of Passau, Germany).

Consequently, the ***first research question***, that was addressed in this work is:

What type of features and which methods can be used to recognize operating modes of electronic devices, that can even be operated on a low power CPU as it was used for the iSensor.

In Section 3.5.3 six power related features were defined and combined with rule classifiers, which are able to describe/differentiate device use-modes with little computational effort. Therefore, signals of eight mainstream kitchen appliances were analyzed. It turns out, that besides **how** devices were used (e.g. different mixing levels), for some devices it could be also recognized **what** they have been used **for** (e.g. mixing something liquid until it goes stiff).

An obvious resulting problem is addressed by the ***second research question***:

How well can use-modes of simultaneously operated devices be recognized by using a single iSensor device?

Section 3.5.3.9 will show, that due to the sensor design, simultaneous device usage and the resulting set of operating modes can be differentiated by a single sensor only to a certain degree.

Finally, the ***third*** and last ***research question*** was:

How well can operating modes of devices be recognized under conditions similar to real-life?

Consequently, a detailed, real-time evaluation of the system was performed in Section 3.5.4. Environments similar to real-life and with multiple users were considered. The achieved classification rates of between 83% and 96% showed, that the *iSensor* in combination with defined features and applied recognition methods is able to solve the considered problem in an unobtrusive, easy to deploy and maintain way.

In summary, the ***key contributions*** of this chapter were:

- An unobtrusive system that is able to recognize **how** and **what** mainstream electronic devices have been used **for** in real-time.
- The definition and use of features as well as rule sets (based on signal analysis), that are able to solve the problem considered and can even be operated on low-power CPUs.
- An evaluation with multiple end-users under conditions close to real-life.

3.4 Design of a Sound-Based Water Flow Measurement System

This section introduces a novel way to approximate the water consumption of common water taps. First, the concept of the system and its architecture is shown. As a next step, a detailed description of applied signal analysis and data processing methods is given. It is followed by the introduction of a concrete system implementation. Finally, the system is evaluated by several real-time and real-world experiments.

3.4.1 System Overview

The system uses a low-quality, low-cost microphone (an off-the-shelf Bluetooth headset). The microphone can easily be attached to standard sized inflow water pipes (e.g. common water taps). In this way, sound generated by running water is recorded and raw audio data is transmitted to a central processing unit via the headset's Bluetooth function. A chain of processing algorithms is used to recognize whether water is running or not. In the case of running water, the water consumption is approximated as well. Figure 3.1 visualizes the principle of the system.

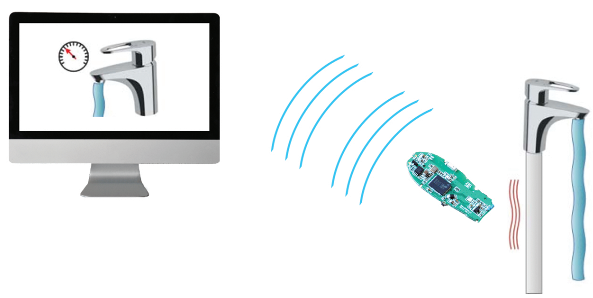


Figure 3.1: System architecture: A Bluetooth headset attached to a water pipe sends raw audio data to a central processing unit.

Audio data was analyzed offline in a first step to find relevant audio features. The resulting feature set was used by machine learning algorithms to distinguish between sound coming from different levels of running water, environmental noise and silence. In order to create an environment-specific model or better said to adapt the system to environmental conditions (training procedure), an initialization step which is easy to perform has to be conducted once. There, sound data of a silent surrounding and six different water flows without noise were recorded (see Table 3.1). In order to make the system resistant against surrounding noise, rule sets were applied on top of recognition outputs delivered by machine learning algorithms. Finally the system's ability to calculate the amount of water consumed was evaluated by comparing three machine learning techniques (varying in complexity) and by adding two different rule sets on top of the best validated machine learning technique.

3.4.2 Algorithms and Methods

This section explains the applied processing tasks of the system proposed in detail. In the following, the data processing chain, the feature selection process and finally a basic evaluation using different machine learning tools and the rule sets on top of them are shown.

3.4.2.1 Data Processing Chain

Figure 3.2 visualizes the corresponding data processing and the data classification chain. As step one to four are basic steps in many sound processing applications, they will be introduced briefly. The focus in this work is on the evaluation of different classifiers and new implemented

Table 3.1: Analyzed water levels and corresponding water flows. The flow value for a constant water level was measured with the help of the flat's main water meter.

Water Level	Flow Value (ml/sec)
0	0
1	7.66
2	34
3	56
4	90
5	125
6	145

rule sets, which are used to filter out false classifications and to improve the accuracy of the water consumption approximation. I would like to note, that the implementation/adaption of "Bluetooth Microphone Reader", "FFT", "Mel Frequency" and "Log10" was done by Alejandro Ibarz (Tecnodiscap, University of Zaragoza, Spain).

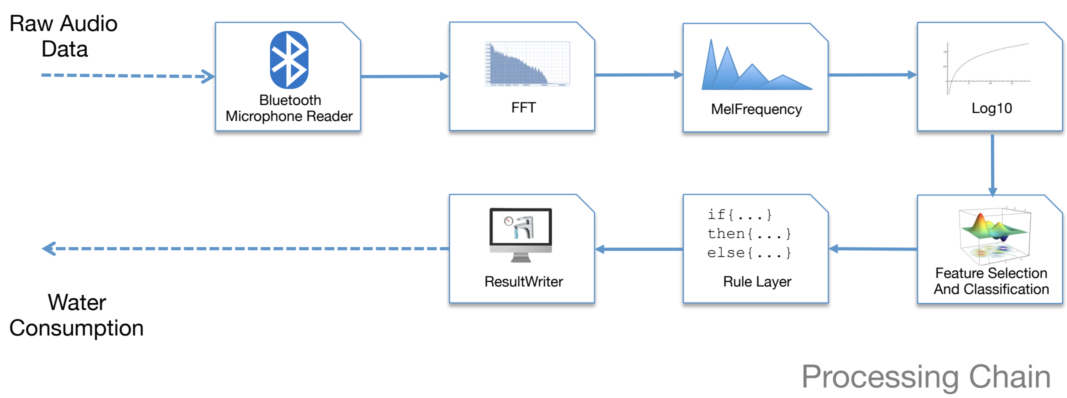


Figure 3.2: Processing chain. Raw audio data is imported and processed step by step. Finally, information about water consumption events is provided in a human readable way.

Bluetooth Microphone Reader: The processing chain starts with a reader task which builds the connection between the processing unit and the Bluetooth microphone. This first task reads raw audio information using synchronous connection-oriented (SCO) Bluetooth logical transport[26]. There, the master is able to create up to three SCO links to different slaves. Consequently, the system is able to get data from three audio devices at the same time. This fact improves the scalability of the system. Nevertheless, this work does not consider multiple sensor devices.

Audio Processing (FFT, Mel Frequency, Log10): The audio processing step is used to extract features from raw audio data. As a first step, raw audio data delivered by the Bluetooth Microphone Reader task is processed using Fast Fourier Transformation (FFT) in order to obtain the spectrum of the current sound signal by using a 8192 point FFT for improved frequency resolution and a rectangular window with no overlapping. Finally, the FFT provided a single sound data spectrum per second. Afterwards, the FFT frequency values obtained are mapped to "mel scale" in order to reduce the size of the spectrum. Therefore a configuration based on 31 triangular overlapping windows, following a logarithmic pattern (see [SJN37]) was chosen in order to obtain a high number of features and to keep a good frequency solution. This procedure

[26] Bluetooth Special Interest Group, https://www.bluetooth.org/apps/content/

delivered a feature vector containing 31 features. Finally, on each feature a logarithmic function to base 10 was performed.

Feature Selection: As the system so far provides a fairly high dimensional feature space which could cause over-fitting and high computation costs, this step performs a dimensional reduction. Therefore, training data for the water flows considered was recorded and calculated features were ranked by weighting methods as Information Gain and Gini Index (see [LM98b] [LM98a]). The seven most valued features were finally chosen as a feature set for the following classification tasks.

Classification: The resulting feature set was used as input for various classifier paradigms. The objective was to compare recognition rates achieved by "simple" classifier paradigms as a Decision Tree (see [Bis06]) or a kNN (k nearest neighbor; see [Bis06]) with more complex classifiers as support vector machines (SVM) (see [CST00]). Classifiers were evaluated using different parameters and detailed evaluation results are shown in Section 3.4.2.2.

Rule Layer: This layer contains several rules which were used to improve the accuracy of the chosen classifiers by reducing the impact of false classifications. As it is almost impossible to record training data for all possible types of surrounding noise, this layer defines rules to filter out sound samples which are quite different to the recorded water flow sounds and hence are assumed to belong to the background noise. The rule layer is applied on top of previously performed classification and data processing tasks. It is also used to approximate the amount of water consumed.

ResultWriter: The last task is used to transform the recognition result in a form which is easy to understand and readable for humans. For example, the amount of water used and the current water flow level can be displayed on mobile or stationary screens.

3.4.2.2 Feature Ranking and Data Classification

It was already mentioned that only a feature subset containing the highest ranked features was chosen for further processing and classification tasks. Consequently, the generalization ability was increased and at the same time computing costs were reduced. Figure 3.3 shows examples for calculated Mel-filter values for different water flow levels as well as for silence.

Mel-filter channels from 6 to 20 might be dedicated to distinguish between different water flows and silence. This assumption was confirmed by feature ranking algorithms (Information Gain and Gini Index) using recorded reference data, which include 100 samples for each water flow considered (see Table 3.1). The seven highest valued channels were channel 7, 8, 9, 10, 11, 14 and 20 (see Table 3.2 for more detailed information about the chosen channels).

Table 3.2: Important feature channels: Frequency information (Hz).

Channel	Start	Center	Stop
7	308	370	437
8	370	437	508
9	437	508	583
10	508	583	663
11	583	663	747
14	837	932	1033
20	1504	1640	1784

Consequently, only these channels were considered for further classification tasks. A 10-fold cross-validation with stratified sampling (see [Bis06]) was applied to pre-recorded training data (100 sound samples for each class) in order to ensure generalization ability. In addition to that, 50 samples (hereafter called $TestSet$) for each class were recorded to evaluate the trained system.

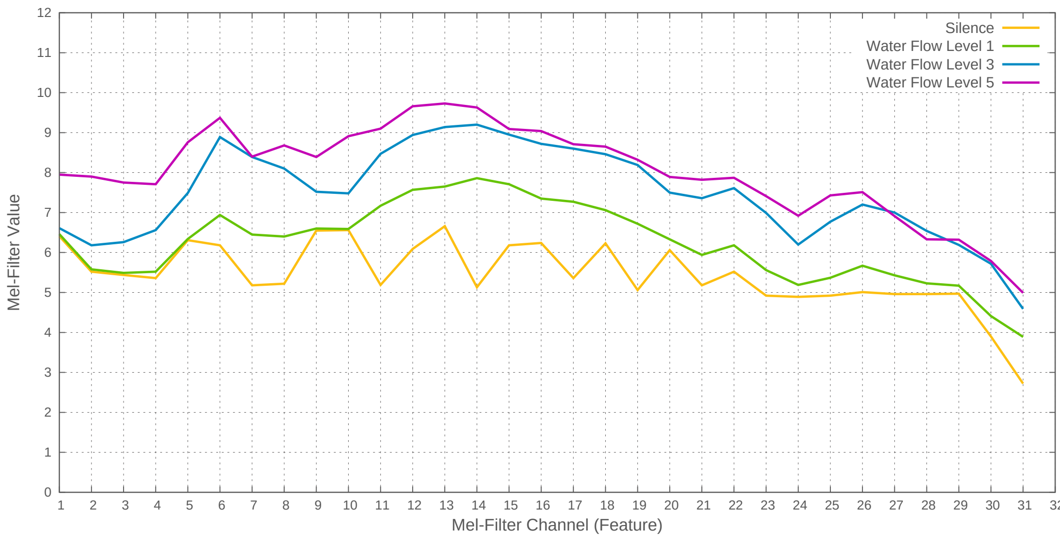


Figure 3.3: Mel-filter channel snapshot for different constant water flows and silence. Experiments showed that Mel-filter channels below 5 and above 22 varied quite strongly even for constant water flows.

The $TestSet$ also includes sound data coming from water levels which are very close to recorded constant flows. In this way, the impact of slightly changed water flows on the classification results was evaluated. Table 3.3 shows the results achieved for each classifier and considered parameters. The evaluation process focuses on the ability to assign recorded sound samples to six different water flow levels and silence.

Table 3.3: Achieved performance for considered classifier paradigms and settings.

Classifier	Settings	Results: 10-fold CV	Results: $TestSet$
Decision Tree[A]	max depth = 10	99.29% +/-0.96	66.11 %
	max depth = 5	97.29% +/-2.07	65.45 %
kNN[B]	k=10	100%	90.70%
	k=5	100%	90.37%
SVM	RBF Kernel, C=0, $\gamma = 1/7$	97.57% +/-0.91	82.72%

[A] The decision tree uses the Information Gain as criterion for numerical splits.
[B] The Euclidean distance was chosen as distance measure for kNN.

Each classifier paradigm was able to provide outstanding recognition rates above 97% using a 10-fold cross validation. But when applying $TestSet$ data to trained classifiers, the performance of each paradigm decreased significantly. Only the kNN classifier was still able to provide a good recognition rate of almost 91%. When analyzing the results in more detail, it can be seen, that all classifiers are able to distinguish between "silence" and water flow levels 1, 2 and 6 very well (about 90%). However, they fail in distinguishing between water flow 3, 4 and 5. The reason for that might be, that sound data for such classes are placed very closely together. Figure 3.4 visualizes the $TestData$ set for channel 7 to 9 and by considering only these channels the previously mentioned assumption can be confirmed. But of course when using all seven channels the feature space could look quite different.

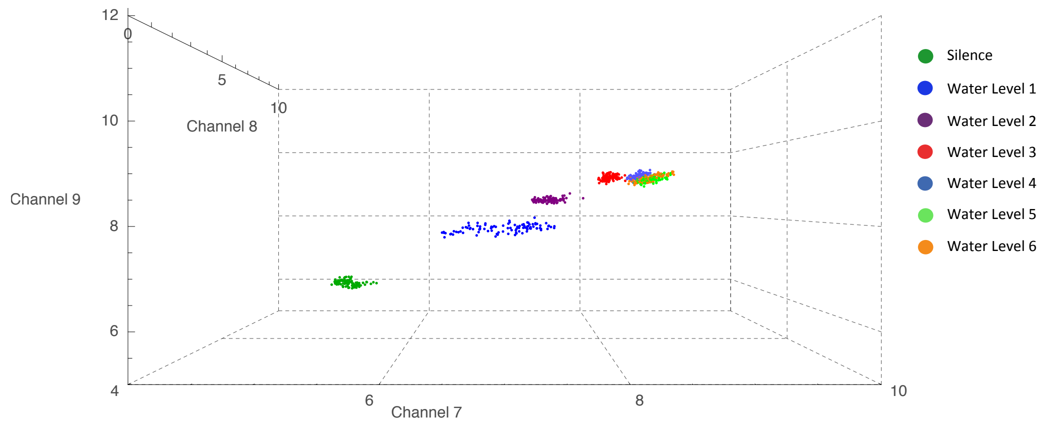


Figure 3.4: Training data visualization using channel 7, channel 8 and channel 9.

The kNN paradigm is chosen for further evaluations as the decision tree as well as the Support Vector Machines show over-fitting and are not able to handle data coming from water flows which are close to reference flows considered.

3.4.2.3 Rule Sets

So far the influence of surrounding noise was completely neglected. As human voices, music or operating sounds of electronic devices are quite usual in real-world environments, rules were defined to recognize and filter out such ambient sounds. In order to recognize background sound, machine learning paradigms require training samples for each class including background data. In real-life applications it is neither comfortable nor possible to record every kind of background noise that could appear. Due to the fact that background samples are missing, noise will be recognized as water flow so far. To avoid this fact, a distance measure was defined which is used to decide whether the classification output provided by kNN was generated by sound coming from water flows or from background noise. Hence, the rule layer tries to filter out false classifications coming from ambient noise and in this way it is used as a noise reduction step. The defined distance measure is shown in equation 3.1.

$$distance = \begin{cases} 0, & \text{if } (kNN_{dist} > 40) \\ 1, & \text{otherwise} \end{cases} \tag{3.1}$$

kNN_{dist} is defined as the sum of distances between the current feature vector containing all 31 features and the k nearest feature vectors (including all 31 features) of the winner class. If the kNN classifies a sound sample as running water (the sample is assigned to one of the water flows considered), the recognition result (specific water flow level) is only accepted if the distance is equal to 1. Otherwise, the current sound sample is assumed to be background noise. The kNN_{dist} threshold is environment independent (this fact is shown in a later section) and was chosen by analyzing experimental results. Although the system is now able to filter out surrounding noise, the fact that very loud noise drowns the running water signal is still a problem when it comes to water consumption calculations. Consequently, a set of rules was defined and used to approximate the water consumption during the existence of ambient sounds. Besides that, rules are also used to reduce false classifications as some noise samples are very similar to water sounds and hence they cannot be rejected by the distance measure applied. The following four rules were defined:

(I) In order to filter out single outliers and consequently reduce the impact of false classifications, a water flow is only recognized if during the last three seconds every sound sample was classified as water flow.

(II) To avoid single false classifications of water flow confusions, the last five recognized water flows are considered and the most present water flow level in this set is chosen as the real water flow.

(III) If a recognized running water flow is interrupted (drowned) by continuous noise lasting not longer than 10 seconds, the system ignores this fact and takes the last recognized water level as the current water flow. If surrounding noise overlaps a present water flow for more than 10 seconds, the system assumes that the water tap was turned off.

(IV) If a water flow is interrupted by "silence", the system assumes that the water tap was shut.

It is clear that some of these assumptions may lead to wrong results. One example could be that the user has forgotten to turn off the water tap and very loud surrounding sounds coming from a radio overlap the sound signal. In this case, the system assumes after 10 seconds that the water tap was turned off, which will lead to a significant error concerning the calculated amount of water consumed.

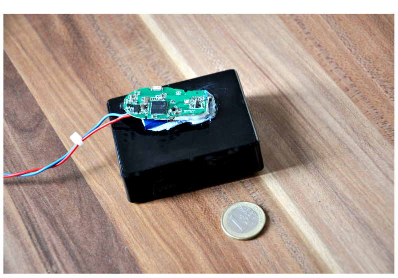
The next section shows a concrete implementation of the proposed system, which was deployed and evaluated in a real-life scenario.

3.4.3 System Implementation

This section shows how the system was realized. In the following, used hardware and software tools are described in detail.

3.4.3.1 Hardware

The system uses a commercial and low-cost off-the-shelf Bluetooth audio device (Headset) as sensor system. The headset's case and speakers were removed and the remaining device was mounted outside a small box (7.5 cm x 5.5 cm x 2.5 cm). The microphone was put inside the box. In order to insulate it from surrounding noise, the box was filled with foam. Figure 3.5 shows the sensor from the inside and outside. Two holes on opposite sides were made, to pass flexible and standard sized inflow water tap pipes through the box. Finally the box cap was closed with four small screws to protect the microphone and to reduce the surrounding noise. Consequently, the sensor can easily be attached to the inflow pipe of a water tap (see Figure 3.6). This fact implicates that neither the pipes nor the water tap itself has to be removed to place the sensor and hence the system is easy to install and due to its small size very unobtrusive. Although the introduced system is battery operated, it can easily be switched to be plugged into standard power sockets. In many cases power sockets are placed nearby water pipes like under the kitchen sink or next to the water taps in bathrooms. Consequently, maintenance effort for the introduced system is not required in practice.



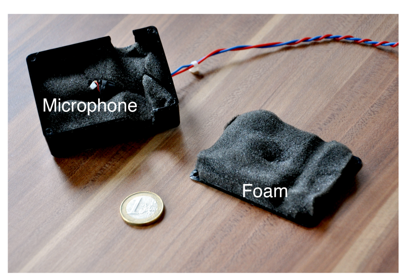


Figure 3.5: Left: Box with a pipe opening and the attached Bluetooth headset device on the top. Right: Box covered inside with foam; The microphone is placed close to the pipe gap.

Figure 3.6: Left and Middle: Sensor mounted on a sink's cold water inflow pipe. Right: Hot and cold water inflow pipes.

3.4.3.2 Software

Due to the fact that the chosen sensor can only be used to digitize and to broadcast sound samples coming from pipes, the online processing of raw audio data was performed on a central computation unit. There, the Context Recognition Network (CRN) Toolbox[27] (see [BAL08]) was used to realize the signal analysis and data processing chain. The CRN Toolbox (CRNT) is a software tool which was designed for POSIX operating systems. It is optimized for the implementation of multi-modal and distributed context recognition systems. The CRNT already contains a broad set of basic algorithms (called tasks) which are often used in signal analysis and data classification applications. Examples are filters, classifiers, feature calculation tasks and common algorithms used in specific applications such as image and audio processing. Apart from that, tasks are stand-alone items providing generic interfaces. Consequently, complex data processing chains can simply be created by interlinking different tasks. The CRN Toolbox was used in this work as it already includes a set of efficient implemented sound processing algorithms (e.g. Fast Fourier Transformation). Besides that, the integration of new and application specific tasks can be done with ease thanks to the concept of generic interfaces. Finally, having a platform independent solution was also one of the main reasons for choosing the CRN Toolbox. Hence, the shown system can be easily ported or installed on smartphones, tablets or personal computer systems.

The following section shows a detailed evaluation of the described system within real-life scenarios.

3.4.4 System Evaluation

The recognition quality of the introduced system is evaluated with several experiments in different locations and scenarios. The objective of all experiments was to validate how well the amount of water used can be approximated using only seven discrete water flows (see Table 3.1). Water measurements, provided by a standard main water meter (as found in common households), were used as ground truths for all experiments. This section evaluates and compares three different system configurations: The raw recognition output of a trained kNN classifier (hereafter kNN_{RAW}), a kNN paradigm combined with the introduced noise reduction layer (hereafter kNN_{NR}) as well as a kNN paradigm in combination with a noise reduction layer and a rule set, which was used to filter out outliers and false classifications (hereafter kNN_{NR+RS}).

[27]The CRN Toolbox is available under LGPL from http://crnt.sf.net.

In the following three experiment scenarios were investigated: First, a detailed experiment was performed to evaluate how well the amount of water consumed could be approximated within a silent as well as a noisy environment. Moreover, the impact of noise coming from nearby water pipes was evaluated. Second, the amount of water used was calculated during a common preparing food event. Finally, the generalization ability of the system was evaluated. It was investigated whether defined core thresholds (noise reduction layer) and parameters (rule set) were also valid on a deviant water tap located in a different environment.

The water consumption was calculated by summarizing current classified water flow estimates over time for each water tap usage event. The system was evaluated in real-time.

3.4.4.1 Approximation of Water Consumption

This section evaluates how well the amount of water consumed can be approximated while considering only six different water flows plus "silence" (see Table 3.1). The sensor was mounted on the cold water inflow pipe of a common water tap located in a real kitchen environment. Based on this setup, three different experiments were performed. First, the amount of water used was calculated in a completely silent environment. The second water flow measurement experiment was performed during the presence of everyday ambient noise. Finally, only environment noise was produced (all in all three minutes of noise sounds) in order to evaluate how strongly the system reacts to such sounds. Everyday sounds, that usually appear in a real-world kitchen environment, were considered as noise. Examples are conversations between people, the opening/closing of doors and cupboards, operating mode sounds of electronic devices such as mixers, microwaves and extractors, listening to music and putting dishes into/removing them from the sink. The main water flow meter of the flat was used as a ground truth for a reliable comparison between the real amount of water used and the system's approximation. Figure 3.7 shows the results achieved for the system configurations considered.

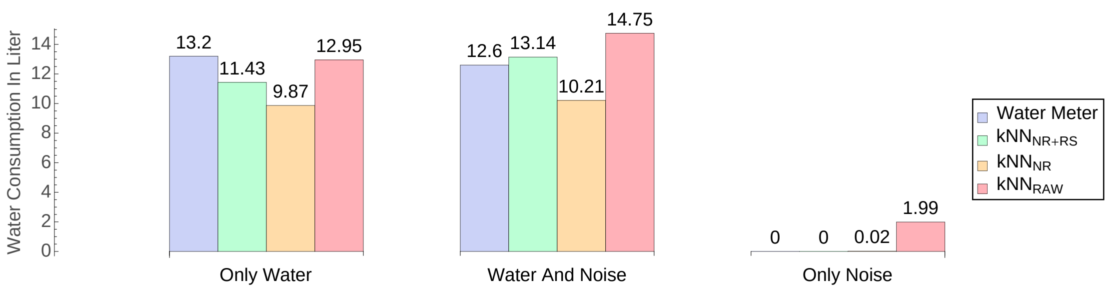


Figure 3.7: Experiment results for various system configurations and different environment conditions.

In environments without any surrounding noise, very good results were achieved by kNN_{RAW}, which uses neither noise nor outlier reduction techniques. There, the system's approximation was only 2% less than the real amount of water consumed. However, when ambient sound is present (as is usually the case) kNN_{RAW} provides much higher water consumption levels due to the fact, that every loud noise sample is falsely assigned to one of the considered water flows. While having running water and noise, 17% more water was calculated as consumed than really used. Even when analyzing the impact of noise samples on the system's recognition quality, it can be seen that kNN_{RAW} wrongly detects water consumption by nearly two liters. This fact shows the importance of a noise reduction layer. When using kNN_{NR} almost no noise samples were falsely classified as running water and consequently only 0.02 liters of consumed water were wrongly calculated. However, this system configuration is not able to approximate the amount of water consumed with the same accuracy as kNN_{RAW}. When having both, environment noise and running water, kNN_{NR} calculates 19% less consumed water than was exactly measured by the flat's water meter. Considering only running water without any environment noise, the system's approximation is even 25% lower. This result was expected as some sound samples of running water were classified as noise and hence running water information got lost. Such problems should be handled by additional rules included in kNN_{NR+RS}. This system configuration is able to filter out all surrounding sounds in the experiment in which noise occurs exclusively. Considering water and noise samples, only 4% more water was calculated than consumed than really used and in the case of having only water sounds 13% less consumed water was approximated. Although the results achieved are quite good, the exact amount of water consumed could not be approximated by any of the introduced systems. The reason for this is the following. On the one hand, kNN_{NR+RS} estimates the current water while having overlapping sound (see rule definition (III)). On the other hand, the system needs some seconds to switch to a new water level after the current water flow has changed (see rule definition (I) and (II)). Consequently, information about the current water flow level get lost. Apart from that, the water consumption calculation is based on only six different water flows, which makes it nearly impossible to calculate the exact amount of water used.

So far, only cold water was measured as the water sensor was fixed to the cold water inflow pipe. As the water sound coming from nearby pipes could influence the system, a detailed evaluation was performed using cold water as well as hot water, which runs through a pipe located next to the monitored one. When using only hot water (hot water was running continuously for three minutes), all system configurations except kNN_{RAW} are quite robust to such sounds (see Figure 3.8). kNN_{RAW} wrongly calculated the water consumption by below one liter while

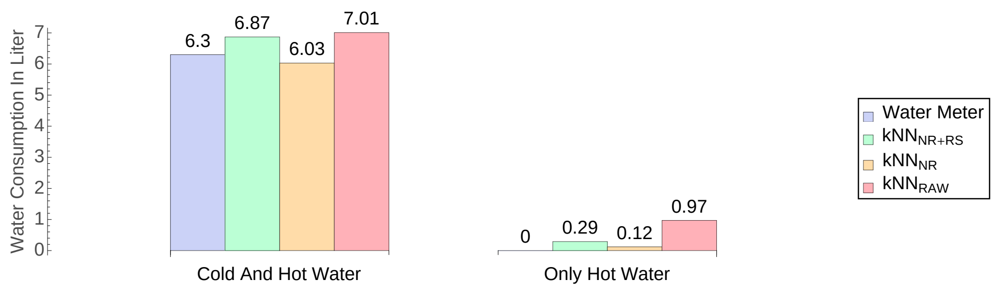


Figure 3.8: Experiment results for hot and cold water flows and various system configurations.

kNN_{NR} and kNN_{NR+RS} provided false approximations of below 0.3 liters. When measuring both cold and hot water, quite good approximations were still achieved by all systems. There the average absolute error of all considered systems and for the amount of approximated cold water used was 8%. This means, that the running water sound coming from pipes located nearby only has a small impact on the recognition quality of the systems introduced.

Table 3.4 shows results which were achieved so far. On average 6.42 liters of water were consumed during all experiments. There kNN_{RAW} as well as kNN_{NR} had a calculation error of 19% and kNN_{NR+RS} of only 10%. Of course the accuracy is not good enough to use the system as an accurate water meter. But for many AAL applications, the approximate amount of used water is still very valuable. Especially in kitchen scenarios the exact amount of water consumption is not really important. Information about what type of pot (large, medium, small) was filled with water can be useful for example to recognize kitchen activities such as preparing food and to distinguish them from other activities (such as washing dishes). The system is used in Chapter 4 to solve a similar recognition problem.

3.4.4.2 Approximation of Water Consumption while Preparing Food

The second experiment was carried out within a real kitchen environment and under real conditions. Therefore the inflow pipe of the kitchen's water tap was equipped with the introduced system. The kitchen is in a shared student flat located in Zaragoza, Spain. Figure 3.9 shows the kitchen environment and the sensor attached to the sink's inflow pipe.

One of the residents was asked to prepare dinner in the same way he is used to doing every day. The person was not informed about how the system works. On the contrary, he was

Table 3.4: Water consumption approximation: Deviation from ground truth in liter.

Scenario	Ground Truth	$\triangle kNN_{RAW}$	$\triangle kNN_{NR}$	$\triangle kNN_{NR+RS}$
Only Water	13.2	-0.25	-3.33	-1.77
Water and Noise	12.6	2.15	-2.39	0.54
Only Noise	0	1.99	0.02	0
Cold and Hot Water	6.3	0.71	-0.27	0.57
Only Hot Water	0	0.97	0.12	0.29
Average Absolute Error	–	1.21	1.23	0.63

Figure 3.9: Left: Kitchen environment. Right: Water sensor (black box) attached to the inflow pipe of the kitchen sink.

only told to prepare food (pasta with ham, tomato sauce, toast with pieces of salmon) and to wash the dishes afterwards. The whole cooking procedure, including washing the dishes, took about 35 minutes. During that time the person produced a large amount of surrounding noise. He opened and closed cupboards, the fridge and drawers several times. Besides that, several kitchen tools such as knives for cutting the ham and salmon were used to prepare the food. Loud ambient sounds were produced by music coming from a kitchen radio. Finally, the person was talking to a friend on the phone during the experiment. After the cooking process, the person put all dishes used into the sink (which also produces significant noise) and washed them. The system was evaluated on its ability to approximate the amount of water consumed. Figure 3.10 shows the achieved results for considered approaches.

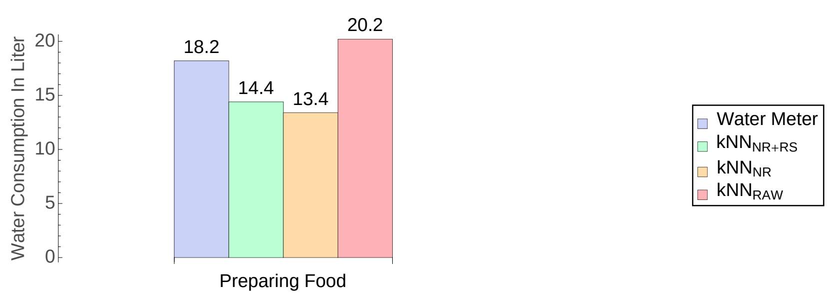


Figure 3.10: Preparing food event: Consumed water and approximated values.

All in all 18.2 liters of water were truly used during food preparation. kNN_{RAW} approximated 20.2 liters of water usage. Again, because of classifying noise as water samples, a higher consumption value was estimated. This outcome was expected due to the results achieved so far. Once again kNN_{NR} calculated lower consumption values compared to the real amount of water used. Only 13.4 liters of water consumed were calculated instead of 18.2 liters. Even kNN_{NR+RS} was not able to produce the same good results as before. This time nearly 21% less consumed water was approximated. The reason for this is, that in such scenarios the water tap is frequently used for very short periods under three seconds (e.g. sketchy cleaning of silverware or washing hands). Precisely such events are suppressed by rule (I). This problem could be solved by reducing the defined threshold of three seconds. However the amount of false classifications will benefit from this modification. Consequently, a trade-off has to be found for each specific application scenario.

3.4.4.3 Universal Usage

Systems introduced so far provided reasonable results when they were trained with environment-specific sound data of different water flows and silence. As the material of inflow pipes could be quite different for various washbowl types, the produced water flow sound might be quite different too. Hence in the following the universal usage of introduced systems is evaluated. This includes the validation of defined rules, selected Mel-filter channels and thresholds adapted to typical kitchen water taps in different locations and on various water tap types. As the system is very easy to adapt to specific environments and different water flow sounds (one-time recording of different water flows and silence), the universal usage of a completely trained system is not analyzed in this work. In this section the system was evaluated in a bathroom and kitchen environment of a typical Spanish residence. Figure 3.11 shows the achieved results.

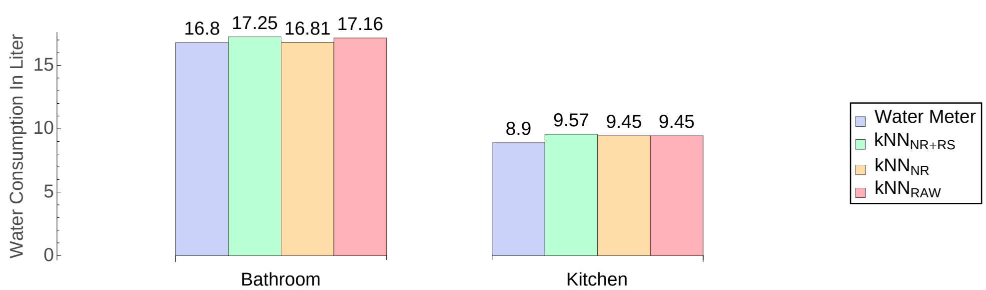


Figure 3.11: Universal usage: Bathroom and kitchen environment without surrounding noise.

It can be seen that by using selected Mel-filter channels, defined thresholds and rules, almost excellent results were reached for different locations without environmental noise. The absolute error produced is below 3% in the case of the bathroom and below 8% in the case of the kitchen experiment. Table 3.5 shows a more detailed error analysis. The maximum average absolute error produced is only 0.5 liters.

Table 3.5: Universal usage: Deviation from ground truth in liters within a kitchen and bathroom environment without surrounding noise.

Scenario	Ground Truth	$\triangle kNN_{RAW}$	$\triangle kNN_{NR}$	$\triangle kNN_{NR+RS}$
Bathroom	16.8	0.36	0.45	0.01
Kitchen	8.9	0.55	0.55	0.67
Average Absolute Error	–	0.46	0.5	0.34

In summary it can be seen, that the core parameters of the system introduced is location independent and the whole system can be adapted to new locations with an initialization process that is easy to perform.

3.4.5 Conclusion

This section introduced an affordable system that is easy to deploy and used to calculate the amount of water consumed. The training procedure which is used to adapt the system to a specific environment is based on simple one-time measurements of different water flows and silence. Although the system needs no training data in terms of environmental noise, it is able to filter out such sounds quite reliably without having a significant influence on the recognition quality. The system was evaluated within different real-life environments and under various conditions including the existence of surrounding noise or noise coming from nearby water pipes. Finally, the system was also evaluated during dinner preparation in the kitchen of a shared student flat. Table 3.6 summarizes the results achieved again. The best results were obtained using kNN_{NR+RS}. This configuration (Fusing kNN with a noise reduction layer and additional rules to reduce the impact of false classifications) creates an absolute error of 10% for the experiments considered. Although the system is still not able to calculate the exact amount of water used, the results achieved are fairly good considering that the approximation is based on a minimal training data set including only six different water flows. Apart from that, the sensor design enables an unobtrusive and easy integration into existing real-world environments even on a large scale.

Table 3.6: System evaluation overview: Deviation from ground truth (in liters).

Scenario	Ground Truth	$\triangle kNN_{RAW}$	$\triangle kNN_{NR}$	$\triangle kNN_{NR+RS}$
Only Water	13.2	-0.25	-3.33	-1.77
Water and Noise	12.6	2.15	-2.39	0.54
Only Noise	0	1.99	0.02	0
Cold and Hot Water	6.3	0.71	-0.27	0.57
Only Hot Water	0	0.97	0.12	0.29
Preparing Food	18.2	2.0	-4.8	-3.8
Bathroom	16.8	0.36	0.45	0.01
Kitchen	8.9	0.55	0.55	0.67
Consumed Water	76	–	–	–
Absolute Error	–	8.98	11.93	7.65
Average Consumed Water	9.5	–	–	–
Average Absolute Error	–	1.12	1.5	0.96

3.5 Operating Mode Recognition of Mainstream Electronic Devices

This section introduces a sensor system that turns common electronic household appliances into smart devices. The system is able to identify the device operated, the current use-mode as well as *what* the device was used *for*. In the following, a general overview about the system is given, followed by a detailed description of the realized power measurement sensor. Afterwards, applied signal analysis and data processing steps are introduced showing that power profile-based rules are sufficient to solve the described problem. Finally, a system evaluation based on eight kitchen devices is shown that was performed by multiple users and under conditions similar to real-life. Moreover, the aspect of operating multiple devices at the same time with one single sensor is examined.

3.5.1 System Overview

We designed and built a wireless sensor device called *iSensor*. The sensor works by using a common power socket. Due to the fact, that it provides the same interface, mainstream electronic devices can be easily attached to the sensor. When using multi-plugs, several devices can even be connected to the same sensor at the same time. The *iSensor* calculates the electric current of the devices connected based on the principle of electromagnetic induction. By using pre-defined and device-specific rules, the sensor is able to identify devices, that are turned on, their operating mode and what they were used for. For example the system is able to detect that a coffee machine is used and to distinguish between corresponding operating modes such as "make espresso" or "make coffee". This information is provided wirelessly using a ZigBee communication module. Figure 3.12 visualizes the concept of the *iSensor* system.

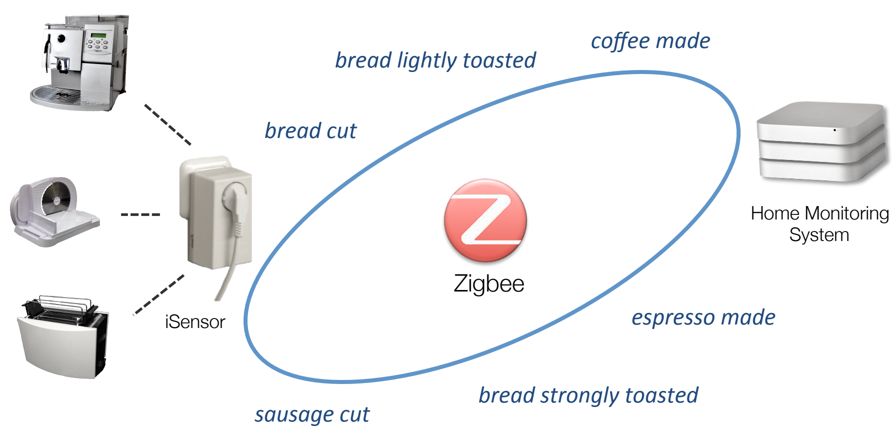


Figure 3.12: *iSensor*: System concept. Multiple devices can be connected to the *iSensor*. The sensor analyzes power features and recognizes devices connected as well as their operating mode. The current use-mode is sent to a central processing unit via ZigBee.

3.5.1.1 Scenarios

The system is evaluated using different mainstream household devices. In general devices can be grouped with regard to their use. Every household contains electric devices for home entertainment (e.g. TV, hi-fi systems, DVD and Blue-Ray players or projectors), kitchen appliances (e.g. toaster, coffee machine or egg boiler) or bathroom appliances (e.g. electric toothbrush, fan or electric shaver). This work considers a set of common kitchen appliances as such devices provide many functions and operating modes (e.g. modes of a common coffee machine: espresso, coffee, heating water and grinding coffee beans). Also, such devices and kitchen scenarios in general are able to deliver very useful information for highly addressed issues, like the monitoring of dietary habits and general stability assessment of daily routines. Table 3.7 lists the appliances considered, corresponding operating modes and what devices were used for in this work. Figure 3.13 and Figure 3.14 show the considered kitchen devices (except the fridge) and how they were connected to the *iSensor*. In order to simulate a realistic kitchen scenario, all devices were connected to the *iSensor* using a multi-plug. Only the fridge was operated by one single *iSensor* as this device is usually not powered through a multi-plug in real-life scenarios.

Table 3.7: Electronic kitchen devices and corresponding use-modes.

Device	Device Usage / Operating Mode
Mixer	Consistency of liquid and mixing level
Juicer	Juice extraction event
Egg Boiler	Amount of boiled eggs and boiling style
Toaster	Lightly or strongly toasted bread
Coffee Machine	Coffee or espresso
Water Boiler	Amount of boiled water
Bread Cutter	Bread cut or sausage cut
Fridge	Door opening events and cooling periods

Figure 3.13: Left: Considered kitchen devices (without fridge). Right: Electronic devices connected to a single *iSensor* using a multi-plug.

3.5.2 Sensor Design

The *iSensor* is based on the principle of electromagnetic induction. The wireless sensor, which is mounted in a small box[28], can be connected to standard power sockets. As the sensor itself is powered by a electrical outlet, there is no need to worry about power consumption which is an important issue in real-life scenarios. The sensor provides the same standard interface as common power sockets and consequently single mainstream electronic devices can be directly connected to the sensor. By using multi-plugs, even multiple devices can be simultaneously connected to a single *iSensor*. This implicates an easy-to-use system which is quite an important fact in real-life scenarios, especially for the elderly. As soon as devices connected are turned on, the sensor measures the induced voltage U_{Ind}. Therefore a current transformer was used which is able to handle devices up to 10 A. Figure 3.15 shows a sensor unit from the outside and the inside.

[28]Height=120mm, width=55mm, depth=65mm.

Figure 3.14: Bread cutter connected to *iSensor*.

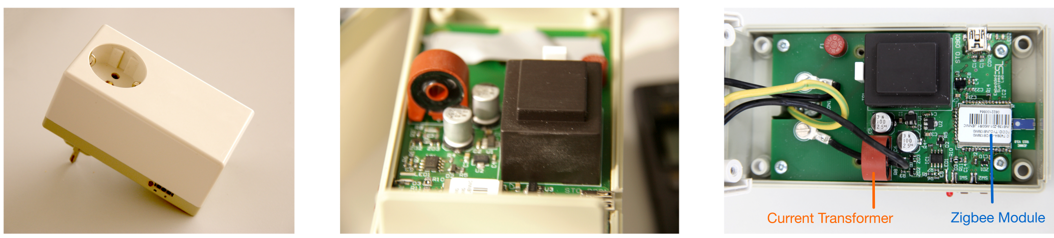


Figure 3.15: Left: *iSensor*; Middle and right: *iSensor* circuit board.

Finally, a 12-bit AD converter (ADC) with a conversion time of 36μs and a feasible voltage range between 0.04 V and 2.4 V digitizes U_{Ind}. The dependency between the induced voltage U_{Ind} and a device's ampere request is visualized in Figure 3.16. The used current transformer

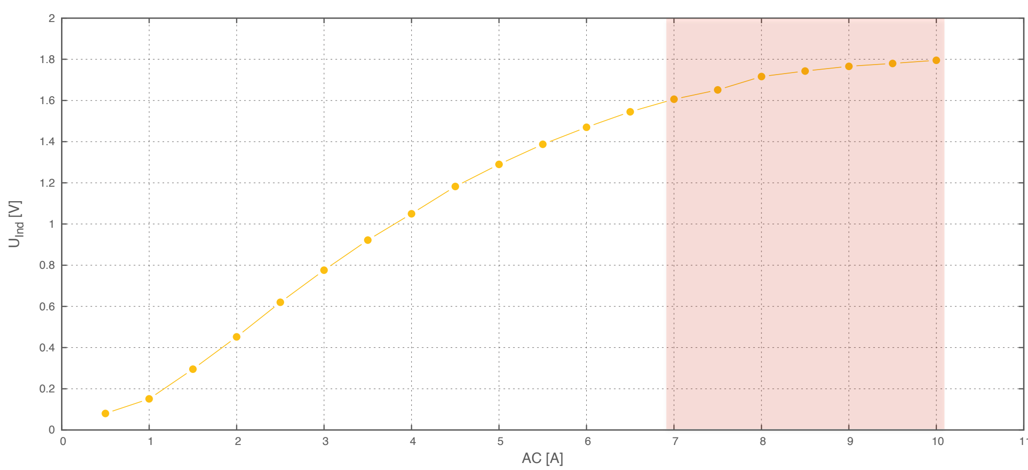


Figure 3.16: Dependency between induced voltage U_{Ind} and a device's ampere request (saturation above 7.0 A).

achieved good results between 1.0 A and 7.0 A. However, the transformer shows saturation for values above 7.0 A. Hence, a current transformer which provides a higher resolution and a larger range is needed when operating devices consuming more than 7.0 A. Especially when using several devices at the same time, values higher than 7.0 A can be easily reached. Consequently, an improved transformer will provide better results, a larger range of applications and reduces costs as the amount of needed sensors can be decreased. Due to the fact, that wired connections between the *iSensor* and data processing units are very obtrusive and not practical in real-life environments, the sensor includes a Jennic Zigbee micro-controller[29] (a low power 32 RISC CPU) to broadcast information to the environment wirelessly. As the micro-controller used is powerful enough to perform easy data computation tasks, the objective was to run processing algorithms on the *iSensor* and to avoid raw data streaming. Consequently, only device events should be transmitted to the environment. The fact that each *iSensor* is a stand-alone system providing information wirelessly and through fixed interfaces, makes the whole system highly scaleable and usable in large-scale applications. I would like to emphasize, that the circuit design and the realization of the *iSensor* was done in co-operation with Karl Stockinger (Embedded Systems Lab, University of Passau, Germany).

3.5.3 Signal Analysis and Data Processing

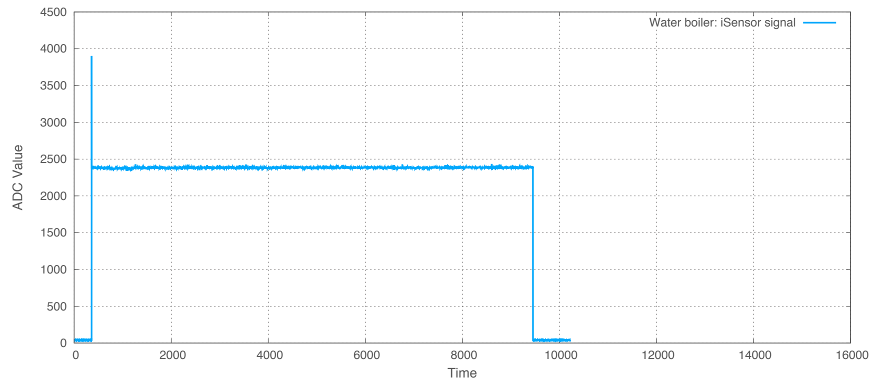
The general idea of this approach is to identify the device, corresponding use modes and what the device was used for by using ADC peak values (ADC_{peak}) as representatives for the current power consumption. ADC peaks were calculated by analyzing the last 600 raw ADC values. Based on the last 21 ADC_{peak} values the following features were calculated: *sum*, *maximum*, *minimum*, *average* and *variance*. The resulting feature vector was used to identify devices and their operating modes. Due to the data processing procedure described, the *iSensor* was able to provide a sampling rate of 2 Hz. One of the considered requirements was that processing algorithms should be designed to run on the integrated Jennic micro-controller and that only operating mode changes are broadcasted. Consequently, this work uses rule sets based on thresholds related to introduced features and time durations of specific operating modes (*duration*). Such features and rules can even be implemented on simple and low-cost micro-controllers as they have been used in this work.

This section will analyze signals coming from kitchen devices when using them in specific operating modes. Based on that fact, rules capturing consumption characteristics were defined for each device and corresponding use modes. In order to identify the device type, a similar approach was applied as described in [PRK+07]. Thus, impulses, which appear when turning on a device, were used to distinguish between the devices considered. The main contribution of this chapter is to recognize the current operating mode and how the device was used after the device type was identified. Therefore, more than 3600 data points were analyzed for each device. In the following, time axes on all figures are related to ADC cycles.

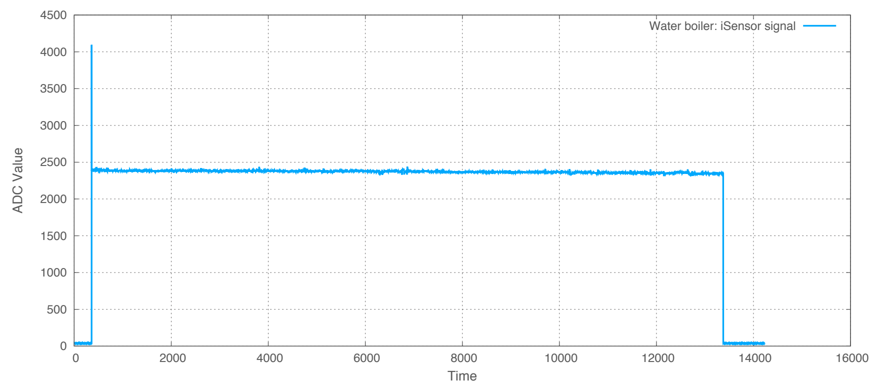
3.5.3.1 Water Boiler

The water boiler in question has a voltage range from 220 V to 240 V and a power rating between 1850 W and 2200 W. After a significant peak when turning on the device, the water boiler provided a constant ADC value of about 2330. Figure 3.17 visualizes ADC values when boiling various amounts of cold water. It can be seen, that delivered values are independent from the amount of boiled cold water. However, the device requires different time intervals to heat various amounts of cold water. In the considered example the water boiler took 469 time samples to heat 1.0 liter and 663 time samples for 1.5 liter of cold water. Consequently, on average 0.002195 liters of water are heated per time sample[30]. Based on the visualized data, rules for distinguishing the water boiler from other devices and calculating the amount of boiled cold water were defined (see Table 3.8).

[29] http://www.jennic.com/ (last accessed on 2013/06/17).
[30] This procedure is only valid for cold water.



(a) Water Boiler: Boiling 1 liter of cold water.



(b) Water Boiler: Boiling 1.5 liter of cold water.

Figure 3.17: Water Boiler: ADC values when boiling different amounts of cold water.

Table 3.8: Rule Set: Water Boiler

Device:	$if \quad ((0 \leq variance \leq 650) \quad and \quad (2300 \leq average \leq 2450))$ $then \qquad device = waterBoiler$
Boiled water:	$duration * 0.002195\frac{liter}{sample} = boiledWater$

3.5.3.2 Fridge

Analyzing ADC values for a common fridge device, door opening events as well as cooling events can be detected. As the latter fact might be obvious (the fridge uses a significant amount of power during cooling), detecting when the door is opened might be more difficult. While the fridge door is open, the internal electric light is turned on. This results in a visible ADC value change. Unfortunately, the constant power consumption of this small light is too low (about 10 W, 240 V) to be measured by the *iSensor* (see Section 3.5.2). However, the *iSensor* is able to detect how long the door was open while cooling as during such periods the power consumption is above the minimum measurable ADC value. Figure 3.18 shows *iSensor* values of a common fridge and corresponding events.

During a cooling procedure the amount of power consumed is almost constant. However, power consumption varies for different cooling procedures. Consequently, the first twenty seconds of a cooling period were used to calculate the current average power consumption $thr_{powerCooling}$. Due to the fact that opening the door is recognized based on $thr_{powerCooling}$, such events cannot be recognized during the initialization phase. Table 3.9 shows derived rules for opening the door and cooling periods. As already mentioned, during cooling periods closing

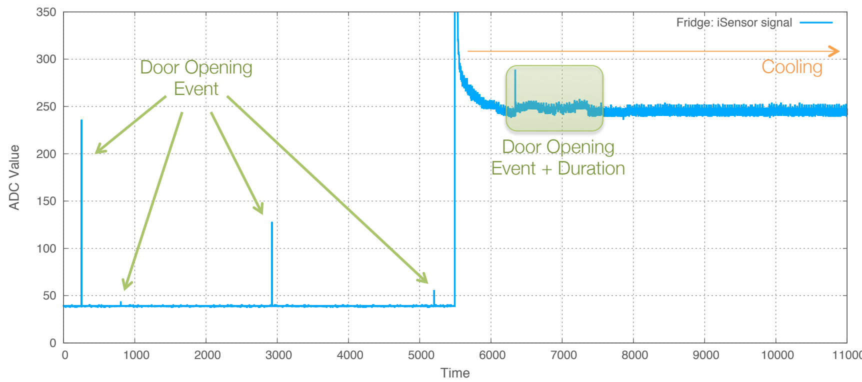


Figure 3.18: Fridge: ADC values for door opening events and the fridge's cooling procedure.

the door can be recognized too.

Table 3.9: Rule Set: Fridge

Activities:	if	$((sum > 842) \wedge (duration == 1))$ $then$ $doorActivity$
	if	$((sum > 842) \wedge (duration > 1))$ $then$ $fridgeCooling$
	if	$(fridgeCooling \wedge ((thr_{powerCooling} - sum) < -70))$ $then$ $doorOpen$ $else$ $doorClosed$

3.5.3.3 Bread Cutter

The bread cutter device evaluated has an operating voltage of 220 V and a rated power of 100 W. The *iSensor* delivers relatively constant and characteristic ADC values in the case of an idle running bread cutter. As slicing food uses much more power, such periods can be recognized reliably. However, the objective was also to evaluate if the sensor is able to recognize what type of food was cut. Therefore, generated ADC values were analyzed for cutting slices of two common foods: bread and salami. Figure 3.19 shows the sensor's output while slicing. Considering both, the duration of a cutting event as well as the maximum ADC value, it is

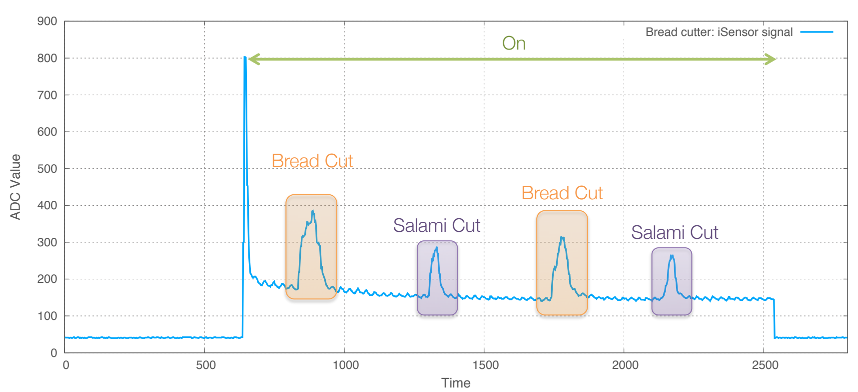


Figure 3.19: Bread Cutter: ADC values for cutting slices of bread and salami.

possible to distinguish between cutting a slice of bread and a slice of salami. Table 3.10 shows the extracted rules to identify the device and to distinguish between running idle and cutting in general, as well as cutting a slice of salami and a slice of bread.

Table 3.10: Rule Set: Bread Cutter

Device:	$if \quad ((0 \leq variance < 4000) \quad \wedge \quad (700 < maximum < 900))$ $then \quad device = breadCutter$
Activities:	$if \quad ((variance < 60) \quad \wedge \quad (100 < average \leq 250)) \quad then \quad on$ $if \quad ((32 < variance < 7000) \wedge (130 < average < 600)) \quad then \quad cutting$ $if \quad ((cuttingDuration < 4) \wedge (maximum < 300)) \quad then \quad salami - cut$ $if \quad ((cuttingDuration \geq 4) \wedge (maximum \geq 300)) \quad then \quad bread - cut$

3.5.3.4 Juicer

The juicer in question has a voltage of 230 V and a rated power of 30 W. Although this work aims at distinguishing between different operating modes of electronic devices, the juicer device focused on can only be used to extract juice. But as such a device is present in many common households, it was considered anyway. Figure 3.20 shows *iSensor* data while extracting juice from oranges.

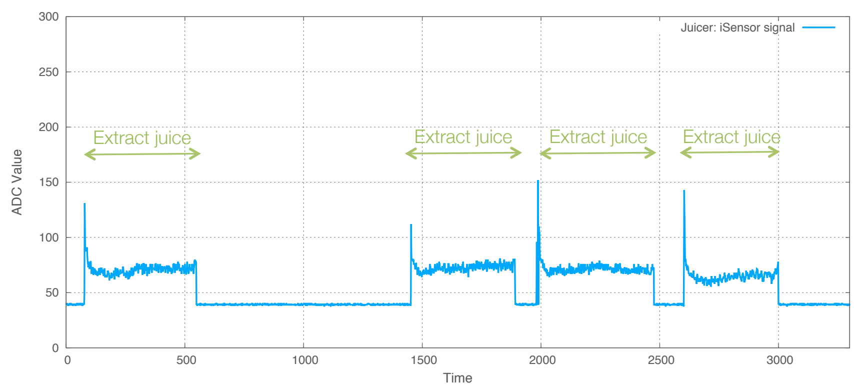


Figure 3.20: Juicer: ADC values for extracting juice from four oranges.

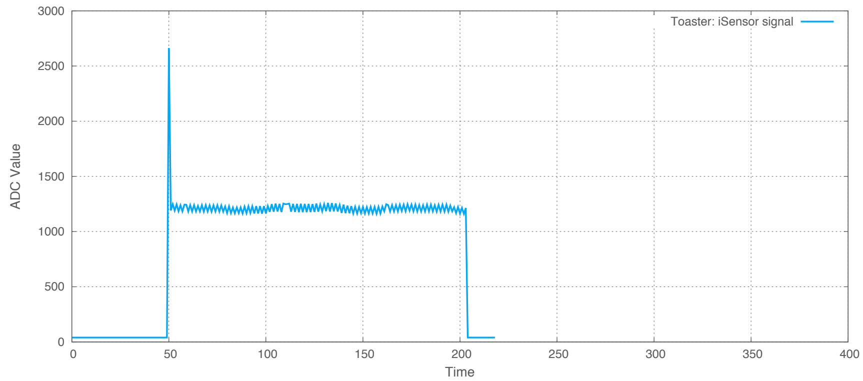
During a juice extraction process the *iSensor* delivers ADC values of about 80. By analyzing ADC characteristics, rules shown in Table 3.11 were defined to identify the device and to recognize single juice extraction events.

Table 3.11: Rule Set: Juicer

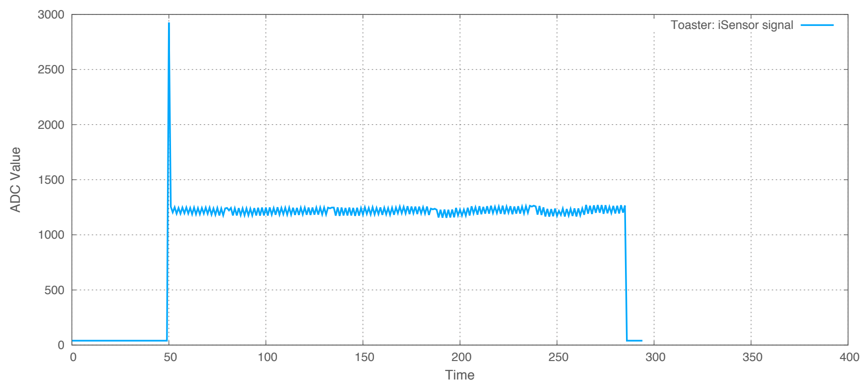
Device:	$if \quad ((0 \leq variance \leq 16) \quad \wedge \quad (50 \leq average \leq 90))$ $then \quad device = juicer$

3.5.3.5 Toaster

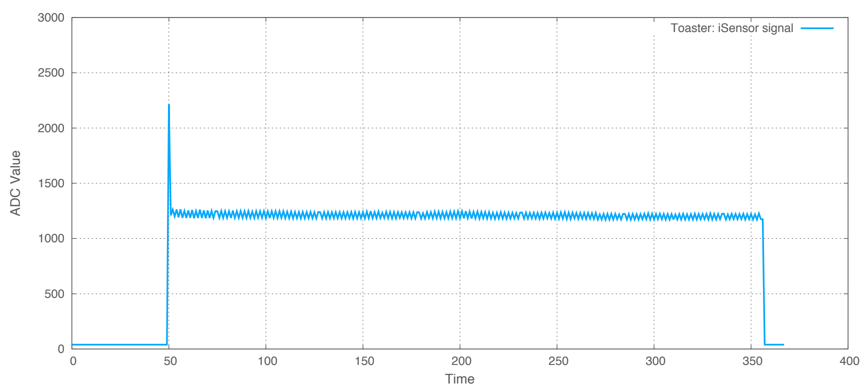
The toaster has a voltage of 230 V and a rated power of 800 W. ADC values delivered by the *iSensor* were used to identify the device, whereas the current operating mode is recognized by analyzing the time duration of a toasting activity. As people have different preferences concerning the bread's toasting level, the objective of this work was to distinguish between three toasting modes: lightly toasted, medium toasted and strongly toasted. Therefore the toaster's operating time range was grouped into three different parts and the center of each group was taken as reference. Figure 3.21 shows ADC values for each of the three toasting levels. As was expected, the *iSensor* delivers quite constant ADC values during the heating process and toasting grade levels can be differentiated by the process's duration. Table 3.12 shows defined rules to identify the toaster device and to recognize the toasting grade.



(a) Toaster: Slightly toasted bread.



(b) Toaster: Medium toasted bread.



(c) Toaster: Strongly toasted bread.

Figure 3.21: Toaster: ADC values when toasting bread lightly, medium and strongly.

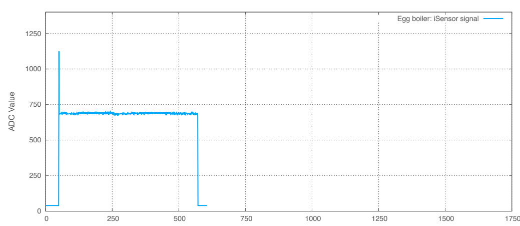
3.5.3.6 Egg Boiler

A common egg boiler with a voltage of 220 V - 240 V and a rated power of 350 W was used. Similar to the toaster device, the egg boiler is identified using ADC values and based on the duration of egg boiling events, different operating modes were distinguished. The egg boiler provides various operating modes so that several eggs can be boiled at the same time and people can choose between multiple types of boiling (from soft-boiled till hard-boiled). Consequently,

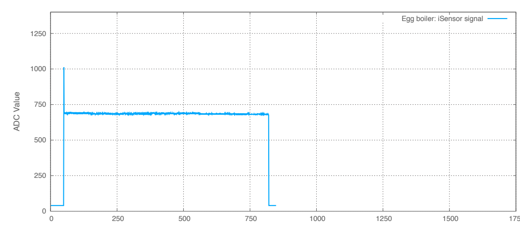
Table 3.12: Rule Set: Toaster

Rule toaster:	if	$((0 \leq variance \leq 600) \wedge (1100 \leq average \leq 1250))$		
	$then$	$device = toaster$		
Activities:	if	$(110 < duration \leq 194)$	$then$	$lightly - toasted$
	if	$(194 < duration < 271)$	$then$	$medium - toasted$
	if	$(271 \leq duration \leq 342)$	$then$	$strongly - toasted$

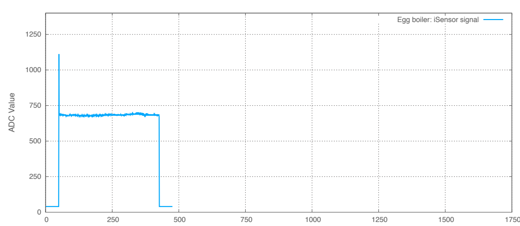
the objective was to distinguish between three levels of boiled eggs: soft, medium and hard boiled eggs. In addition to that, it was analyzed how many eggs were boiled. In the following, the focus was on the differentiation between three and five boiled eggs. Figure 3.22 shows the *iSensor*'s ADC values for different combinations of boiling styles and the number of boiled eggs. Due to the fact that ADC values are quite constant, the combinations considered can be differentiated by analyzing the duration of a boiling event. Boiling five eggs at a time takes less time than boiling three eggs in the same way. This fact is due to the increased egg surface size and the amount of condensed water. In Table 3.13 defined rules to identify the egg boiler and to recognize both the amount of eggs boiled as well as their boiling styles are shown.



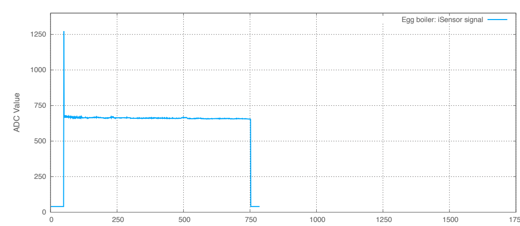
(a) Egg Boiler: Three eggs soft boiled.



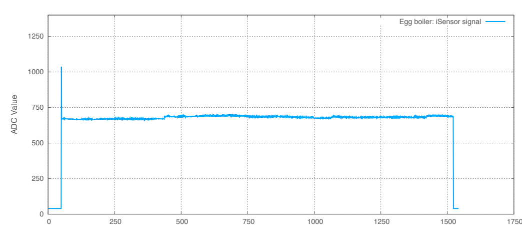
(b) Egg Boiler: Three eggs medium boiled.



(c) Egg Boiler: Five eggs soft boiled.



(d) Egg Boiler: Five eggs medium boiled.



(e) Egg Boiler: Five eggs hard boiled.

Figure 3.22: Egg Boiler: ADC values when boiling eggs (soft, medium and hard).

3.5.3.7 Coffee Machine

Coffee machines are present in almost every household. Thus, the *iSensor* was used to analyze ADC values coming from a mainstream coffee machine with a voltage of 220 V - 230 V and

Table 3.13: Rule Set: Egg Boiler

Rule egg boiler:	$if \quad ((0 \leq variance < 15) \wedge (600 < average < 700))$
	$then \quad device = eggBoiler$
Activities:	$if \quad (450 < duration < 650) \quad then \quad 3eggs - soft$
	$if \quad (750 < duration < 850) \quad then \quad 3eggs - medium$
	$if \quad (2000 \leq duration \leq 2200) \quad then \quad 3eggs - hard$
	$if \quad (250 < duration < 450) \quad then \quad 5eggs - soft$
	$if \quad (650 < duration < 750) \quad then \quad 5eggs - medium$
	$if \quad (1300 \leq duration \leq 1600) \quad then \quad 5eggs - hard$

a power rate of 1450 W. A coffee machine provides different operating modes beside brewing coffee, such as heating water or grinding coffee beans. The modes mentioned later on are not very useful for many applications. In contrast to that, information about the amount of coffee brewed per person and per day is very valuable for health care scenarios. Hence this section aims at recognizing coffee brewing activities. Besides, based on such events the system distinguishes between two different types of coffee brewed: Espresso and normal coffee. Figure 3.23 shows the ADC measurements received for both operating modes. The resulting rule set is shown in Table 3.14.

Table 3.14: Rule Set: Coffee Machine

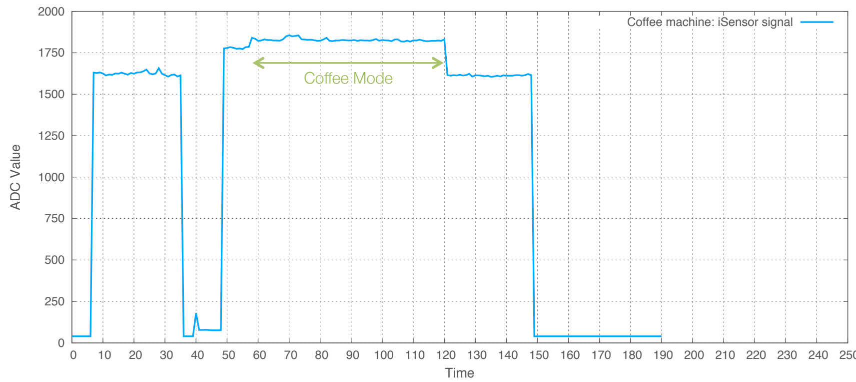
Rule coffee machine:	$if \quad ((0 \leq variance < 1000) \wedge (1700 \leq average \leq 1890))$
	$then \quad device = coffeeMachine$
Activities:	$if \quad (25 < duration \leq 58) \quad then \quad espresso$
	$if \quad (58 < duration < 200) \quad then \quad coffee$

3.5.3.8 Mixer

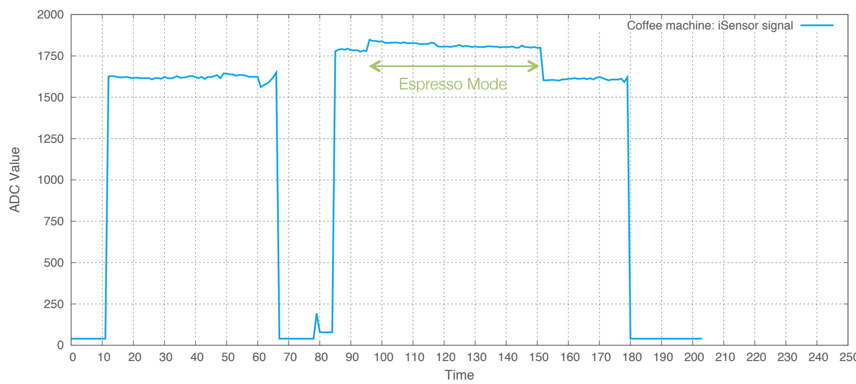
Finally, a mixer device was investigated. The device chosen has a voltage of 220 V - 240 V and a power rate of 100 W. Besides the distinction between two different mixing levels, the objective was to recognize the consistency of the mixed substance. In this work substances with a liquid, medium and creamy consistency have been focused on. Based on such information, AAL services can be created that are able to tell the user the current state of the mixed substance or the device itself can stop the mixing procedure if the desired consistency is reached. Blind people especially could benefit from such systems. Figure 3.24 shows *iSensor* values for different mixing levels and different consistencies of the mixed substance. Corresponding rules are shown in Table 3.15.

Table 3.15: Rule Set: Mixer

Device:	$if \quad ((0 \leq variance \leq 30) \quad \wedge \quad (78 \leq average \leq 150))$
	$then \quad device = mixer$
Activities:	$if \quad (78 \leq average \leq 88) \quad then \quad mixer \quad level1$
	$if \quad (89 \leq average \leq 110) \quad then \quad mixer \quad level2 - liquid$
	$if \quad (110 < average \leq 125) \quad then \quad mixer \quad level2 - medium$
	$if \quad (125 < average \leq 150) \quad then \quad mixer \quad level2 - creamy$

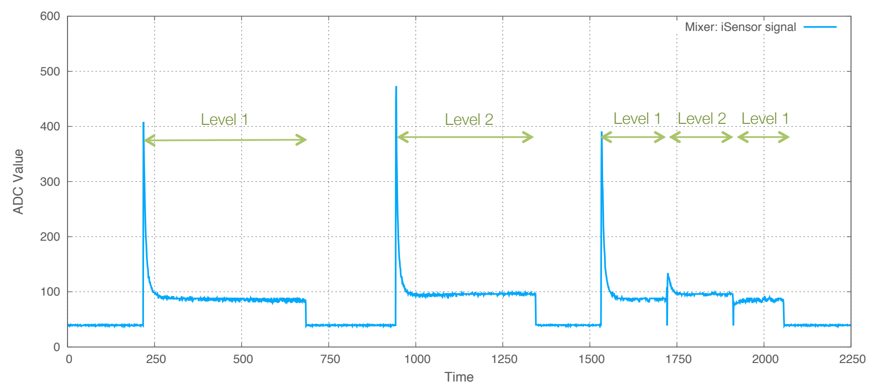


(a) Coffee Machine: Coffee mode.

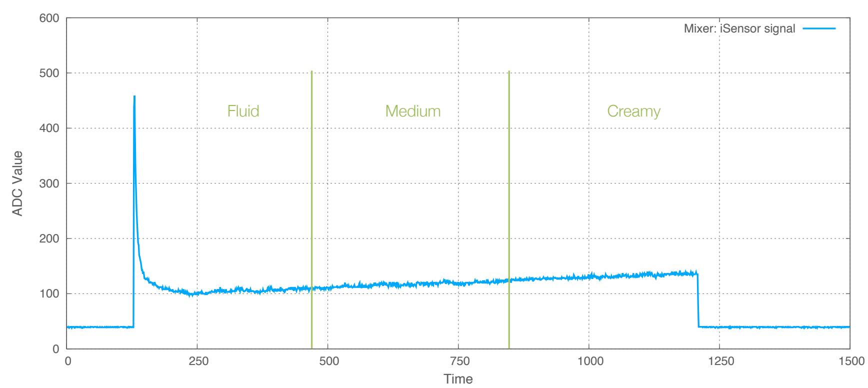


(b) Coffee Machine: Espresso mode.

Figure 3.23: Coffee Machine: ADC values while making coffee and espresso.



(a) Mixer off and on (empty tank – mixing level 1; level 2; level 1-2-1).



(b) Mixer (level 2) when mixing a protein shake (consistency from fluid over medium to creamy).

Figure 3.24: Mixer: ADC values for different operating modes.

3.5.3.9 Simultaneous Device Usage

So far it was shown that the devices considered and their corresponding operating modes have characteristic current signatures, that allow us to identify the device and the chosen operating mode. However, only one device was operated at a time. In a real kitchen environment, this assumption is not realistic, as usually more than one device is being used at the same time. In order to reduce the number of sensors required and consequently costs, this section evaluates whether rule sets can be defined when several devices are connected to a single *iSensor* (e.g. using a common multi-plug) at the same time. Figure 3.25 and Figure 3.26 show *iSensor* values for several devices operated simultaneously. It can be seen, that signal changes coming from

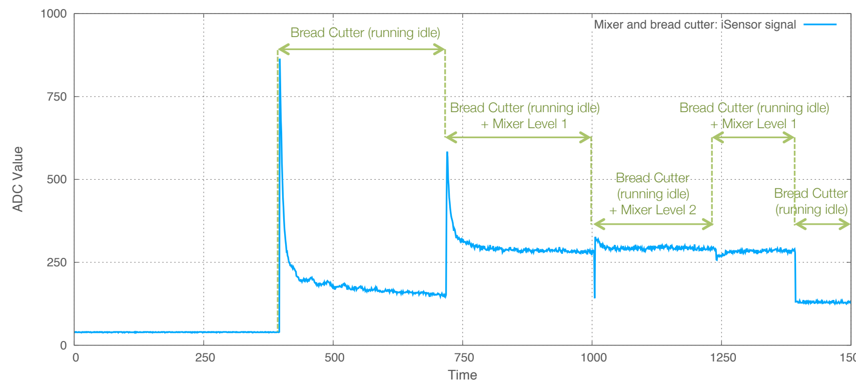


Figure 3.25: ADC peaks for simultaneously used devices: Mixer (level 1 and 2) and Bread Cutter (running idle).

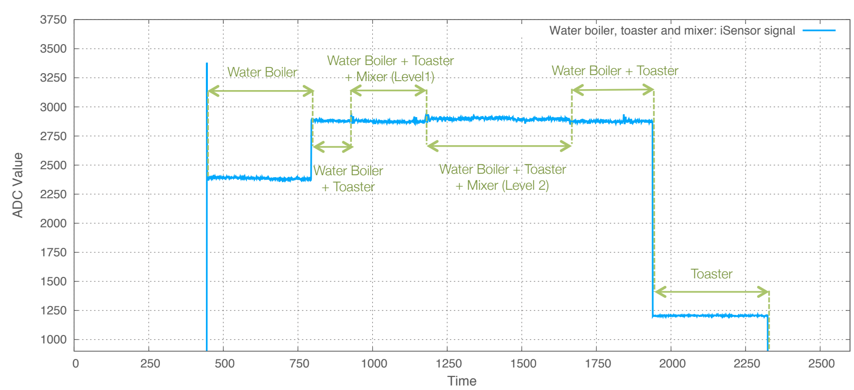


Figure 3.26: ADC peaks for simultaneously used devices: Water Boiler, Toaster and Mixer (level 1 and 2).

newly turned on devices and operating modes can be recognized by the *iSensor* (see Figure 3.25). However, due to the fact that the current version of the *iSensor* shows saturation for higher ampere requests, it is difficult to handle device combinations with high power consumption (see Figure 3.26). As a result the current version of the introduced *iSensor* has to be adapted to handle this problem and to be able to deliver significant signal changes for high power consumption scenarios.

3.5.4 System Evaluation

This section describes the evaluation procedure. All kitchen appliances except the fridge were connected to a single *iSensor* by a multi-plug. Using a similar approach as described in [PRK+07], about 90% of devices observed were identified whilst they were turned on. This work focuses on the recognition of specific operating modes and what the device was used for. Consequently, the following evaluations are based on this issue and show how accurately the introduced system can distinguish between various operating modes after a device was correctly recognized. The following evaluation process was performed online by real end-users at several real-life locations[31]. The little amount of repetitions for each experiment might seem strange at first glance. However, due to the fact that the power consumption of a specific device, as well as corresponding operating modes are quite constant and that food should not be wasted, the considered runs are sufficient to get reliable evaluations. During the experiments 160 pieces of foods were used and the fridge was monitored for 35 hours.

3.5.4.1 Evaluation: Mixer

Apart from distinguishing between different mixing levels, the *iSensor* was also able to recognize the consistency of the substance mixed. In this work the focus was on recognizing whether something liquid or creamy was being mixed. Seven people were asked to prepare chocolate milk (liquid consistency), which consists of 250ml milk and two spoons of chocolate powder. Additionally, every person had to prepare a creamy quark-yogurt dish (250g quark mixed with 100g yogurt). Both mixing procedures were done on mixing level two. As a result 100% of the chocolate milks prepared were correctly recognized as liquid substances and 100% of the quark-yogurt dishes were correctly classified as creamy substances. For 57% of the prepared quark-yogurt dishes a consistency change from medium to creamy was recognized too.

3.5.4.2 Evaluation: Juicer

Six people were asked to extract juice from two orange halves. Consequently, 12 juice extractions were performed. The *iSensor* was able to recognize 92% of the juice extractions performed. Furthermore, no false classifications occurred. Table 3.16 shows the results.

Table 3.16: Evaluation: Juicer

performed	recognized	classification rate
12	11	92 %

3.5.4.3 Evaluation: Egg Boiler

The evaluation includes six people who were asked to prepare different amounts of boiled eggs. In more detail, three and then five soft boiled eggs have been considered. Table 3.13 shows, that thresholds for these combinations are next to each other and hence confusions are very likely. However, the *iSensor* system was able to recognize the two different ways of boiling eggs with an accuracy of 83%. Table 3.17 shows the corresponding confusion matrix.

Table 3.17: Evaluation: Egg boiler

real / recognized	5 eggs soft boiled	3 eggs soft boiled	classification rate
5 eggs soft boiled	6	0	100 %
3 eggs soft boiled	2	4	67 %
average			84 %

[31] During the evaluation procedure, raw ADC peaks were streamed to a processing unit where introduced rules were applied.

One can see, that only two "three soft boiled eggs" events were falsely recognized as "five soft boiled eggs". The reason for this might be that people were not very accurate in adding the exact amount of proposed water, which is of course a basic precondition. In contrast to that 100% of "five soft boiled eggs" were correctly classified. Results for combinations including hard and medium boiled eggs are expected to be even better considering the range of defined rules. Only the most difficult case was evaluated here due to the large amount of eggs needed for further evaluations.

3.5.4.4 Evaluation: Toaster

The *iSensor* was used to recognize the way in which a slice of bread was toasted. Three different toasting styles were considered: lightly, medium and strongly toasted. Six people were asked to toast bread. Each person had to toast two slices of bread: one lightly and one strongly. Based on this information participants had to configure the toaster to reach the desired toasting style themselves. Table 3.18 shows the achieved results.

Table 3.18: Evaluation: Toaster

real / recognized	lightly toasted	strongly toasted	classification rate
lightly toasted	5	1	83 %
strongly toasted	1	5	83 %
average			83 %

All in all the system was able to classify 83% of prepared toast styles correctly. Only one lightly toasted bread was confused with a strongly toasted one and vice versa. This could be due to the fact that the definition of lightly and strongly toasted bread was not specified and participants were not familiar with the toaster and its configurations.

3.5.4.5 Evaluation: Coffee Machine

Six people were asked to prepare two different types of coffee: Espresso and normal coffee. Table 3.19 shows, that the *iSensor* was able to distinguish between these two coffee types with a 92% accuracy.

Table 3.19: Evaluation: Coffee Machine

real / recognized	coffee	espresso	classification rate
coffee	6	0	100 %
espresso	1	5	83 %
average			92 %

During the experiments the following observation was made: Every time the coffee machine brewed a cup of coffee, the water tank was automatically heated. During this time, the device consumes a similar amount of power as when making coffee. Due to the fact that the variance is much higher, such water heating events can easily be filtered out. However, every time the coffee machine is turned on for the first time and consequently both water and the heating elements are cold, the variance of heating water is also the same as for brewing coffee. Due to this fact, the *iSensor* cannot filter out such heating events and brewing coffee is falsely recognized each time the device is turned on. Nevertheless, as this is a static behavior such false classifications can also be filtered out using additional rules.

3.5.4.6 Evaluation: Water Boiler

As was already stated the *iSensor* is able to approximate the amount of boiled cold water. During the experiment different amounts of water ranging from 0.75 l to 1.7 l were evaluated.

In this way the whole capacity range of the water boiler used was covered. Figure 3.27 shows rounded approximation values for five different amounts of boiled water.

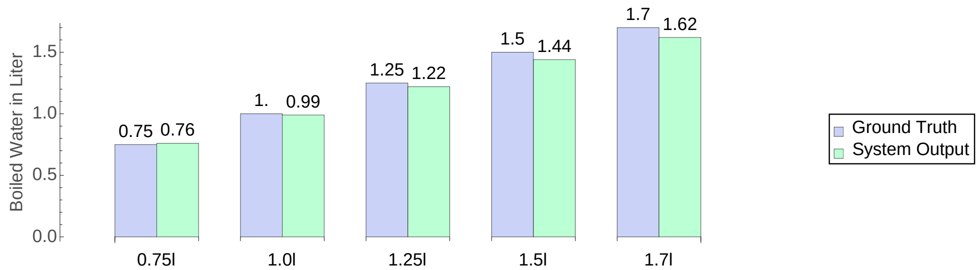


Figure 3.27: Water Boiler: Approximated amount of water boiled compared to ground truth.

The absolute deviation achieved was between 0.006 liter and 0.076 liter. The average absolute deviation was 0.038 liter. Consequently, the system is able to approximate the amount of boiled cold water quite well. A deeper analysis was done in order to investigate to what degree results vary between different approximation steps. So 1.3 liters of cold water were boiled five times one after another. Table 3.20 shows the results achieved. The *iSensor* provides quite constant results for each approximation step. On average 1.262 liters of water was approximated for 1.3 liters of boiled water. The average absolute deviation is 0.038 liters with a pretty low variance of 0.0000547.

Table 3.20: Water Boiler Evaluation: Boiling 1.3 liter of cold water. Approximation and absolute deviation in liters.

Real Boiled Water	Approximation	Absolute Deviation
1.3	1.269	0.031
	1.256	0.044
	1.253	0.047
	1.269	0.031
	1.264	0.036
Average	1.262	0.038

3.5.4.7 Evaluation: Bread Cutter

The bread cutting device was evaluated by six people with the objective of recognizing what type of food was being cut. Thus, people were asked to cut slices of bread as well as slices of salami. All in all 28 slices of bread and 29 slices of salami were cut during the experiment. Figure 3.28 visualizes the results achieved. The *iSensor* was able to provide good classification rates for both foods. 96% of performed bread cuts and 93% of performed salami cuts were recognized. However, false classifications and insertions appeared while cutting salami slices. The achieved recall with respect to salami slices is 93% with a corresponding precision of 90%. In the case of bread slices, a recall of 96% with a corresponding precision of 100% could be reached. The reason for this is, that some people had problems using the bread cutting device and pieces of bread or salami got caught between the blades. This fact led to false classifications as such problems are not covered by corresponding rules.

3.5.4.8 Evaluation: Fridge

The fridge was evaluated in a real office scenario. Therefore a common fridge was monitored in a kitchen environment, which belongs to a research group of the University of Passau. The fridge was connected to a single *iSensor* as such devices are normally placed isolated. The device was monitored for about 35 hours. The objective was to spot and count how many times the door

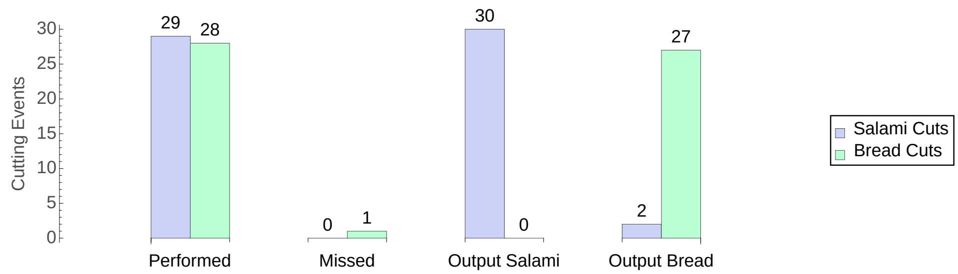


Figure 3.28: Bread Cutter: Evaluation results for cutting slices of salami and slices of bread.

was opened. Every person had to write down the amount of times the door was opened during the experiment. This information was used as ground truth. Figure 3.29 shows the results achieved.

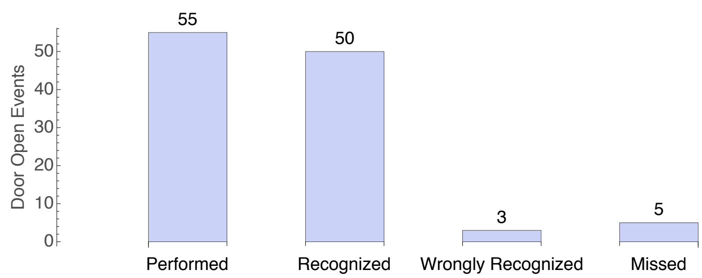


Figure 3.29: Fridge: Evaluation results during a 36 hour observation.

The *iSensor* was able to recognize nearly 91% of the time the door was opened. As the system is not able to detect the door opening during the initialization phase of each cooling procedure, such events cannot be recognized. During the evaluation procedure, 9% of door opening events performed were not recognized. Besides, three door events have been recognized twice. All in all the system was able to provide a recall of 91% and a corresponding precision of 94%.

3.5.5 Conclusion

This section introduced a novel approach based on an inductive sensor to turn mainstream electric household devices into smart devices. The *iSensor* offers many advantages for large-scale scenarios. Among other things the wireless sensor is easy and unobtrusive to deploy (it is simply connected to common power sockets), easy to set up (it is based on a minimal set of reference data), easy to maintain (the sensor is powered through the connected current) and affordable (the prototype costs less than 300€[32]). The sensor was evaluated in real-world environments and with several common kitchen devices. It was shown, that mainstream appliances such as mixers, bread slicers or toasters are able to provide much more information, like "in usage" or "idle" as was investigated by many related approaches so far. Besides identifying devices in use, the system is also able to recognize the current operating mode of electronic devices (e.g. "mixer: level 1" or "mixer: level 2") or even how/what the device was used for (e.g. "3 eggs – soft boiled" or "bread – strongly toasted"). Classification rates of between 83% and 96% and high precision values showed that the *iSensor* and the rule based approach introduced are able to solve the considered problem reliably.

[32]Chapter 4 uses the same recognition principle but based on a similar and commercial sensor for less than 130€. This sensor did not exist at the time the *iSensor* was developed.

However, the proposed system has to face the following problems:

- The system is not able to handle high voltage current as is usually the case for common stoves in many kitchen scenarios. However the monitoring of oven operating modes may strongly contribute to activity detection systems for many applications like cooking. This disadvantage is weakened by the fact, that smart ovens providing detailed information about their use-mode by web services are already on the end-user market (e.g. Miele@Home[33]) in contrast to kitchen appliances focused on in this work.

- The system will fail, if two different devices showing almost the same power profile are used at the same time and are connected to the same sensor. A possible solution could be to include further power features, that may be more device specific such as reactive power or phase angle. Apart from that, the system can be fused with other methods to identify the device. For example, user-device proximity detection systems can be used to recognize the device operated and based on that information, the device use-mode can be determined by analyzing power features.

- The system shown works for wired electronic devices only. However, in common real-life scenarios, electrical appliances are present that can be operated with batteries (e.g. electric shaver or tooth-brush). For such devices, basic use-mode information can be derived only by monitoring their charger devices.

3.6 Conclusion

This chapter introduced affordable, easy to deploy and maintain approaches to turn common household appliances into smart devices. Clearly the monitoring of water taps and the recognition of electronic device usage is not an unnoticed problem. However, this chapter extends the current state-of-the-art approaches as it shows that:

1. Even low-cost sensor systems are able to provide much more information than identifying the kind of device the user is interacting with. It is shown that besides the detection of flowing water, the water consumption can be approximated and besides identifying electronic devices in usage, detailed information about *operating modes* or even *what* devices have been used *for* can be provided.

2. Such systems can be realized based on simple and fast one-time measurements instead of using large training data sets.

3. Such systems can be designed to be unobtrusive and to be usable in real-life and large-scale scenarios.

The resulting advantage is the following. So far almost all mainstream household devices, except for some high-end products (e.g. the smart oven from Miele), are closed systems which are not able to provide their current operating mode in a standardized way. But such information is exactly the basis for high-level activity and context recognition services. For example, smart home applications benefit greatly from intelligent household devices by fusing them with activity recognition systems. Self-learning and self-adapting systems can recognize user habits and significant event-related deviations. Moreover, health care services can support elderly and disabled people to handle their daily life. As it is very unlikely that all household devices will be replaced by their intelligent followers within the foreseeable future, smart solutions must be found to turn common household devices into smart devices. This chapter introduced two different systems working towards this problem.

The assumption that smart devices can significantly improve existing context and activity recognition systems, is confirmed in the next chapter (Chapter 4). There, both approaches were integrated into an office environment. It is shown, that the combination of smart devices with

[33] http://www.miele-at-home.de/de/aktion/mieleathome/656.htm (last accessed on 2013/09/15)

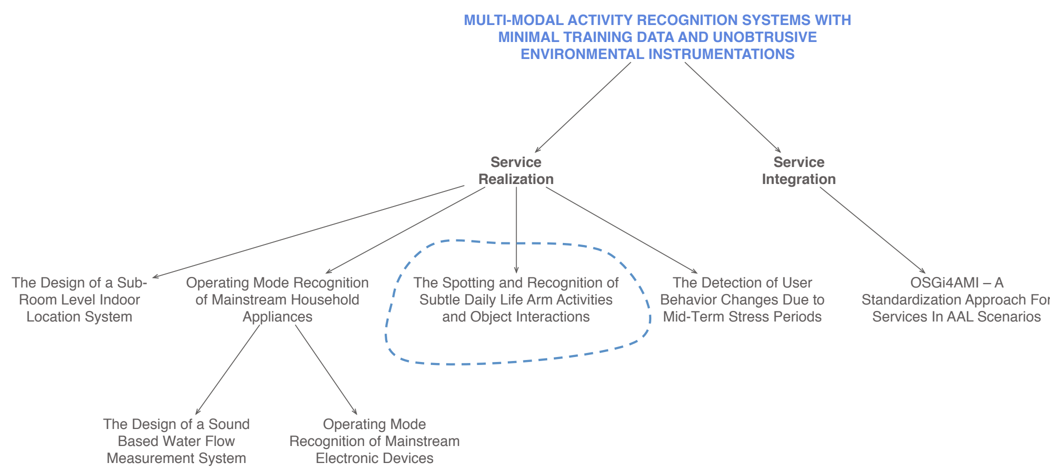
common sensor systems will result in a significant classification improvement. Hence unobtrusive, easy to deploy systems, which are able to make mainstream household devices smart (as introduced in this section), are excellent components for the integration of daily-life activities into real-world context recognition systems – even on a large scale.

4

Spotting and Recognition of Subtle Daily Life Arm Activities and Object Interactions

This chapter is based on my work published in

Gerald Bauer, Ulf Blanke, Bernt Schiele, and Paul Lukowicz. *User Independent, Multi-Modal Spotting of Subtle Arm Actions with Minimal Training Data*, 10th IEEE Workshop on Context Modeling and Reasoning 2013, San Diego.



4.1 Introduction

So far, the thesis has introduced novel approaches related to indoor positioning as well as the use-mode recognition of mainstream household appliances in real-life and large-scale scenarios. This section focuses on the spotting and recognition of hand activities, which is another frequently addressed problem in pervasive computing applications. Amongst others, the information about concrete hand actions is a basic component for the spotting of high-level, complex user activities. Detected hand activities such as "screwing", "hammering" or "sawing" can be used to recognize complex events like "assemble furniture" or even to monitor the progress of such procedures.

Activity spotting in general aims to detect specific actions in a continuous data stream that contains arbitrary activities. Activities of interest are often embedded in a large amount of background data which is called the "NULL" class. In many real-life, large-scale scenarios it is impractical to build reliable models for a class that contains "everything else" (i.e. not belonging to activities of interest). Due to this fact, the spotting of hand activities is a difficult problem to solve. This work makes that problem even more challenging as

- It considers a particularly difficult version of the spotting problem: Only subtle and barely distinguishable actions that are determined by simple and short hand or arm gestures have been focused on. Examples are pressing buttons on various devices, opening/closing cupboards, picking something up or putting it away. Consequently, the NULL class consists of "all the other arm motions that a person may perform" – including motions that are very similar to the relevant actions.
- Subtle arm actions are embedded in a large amount of background data. The ratio between activity of interest and background activity samples is 1:10.
- It faces the problem of collecting minimal but sufficient training data to design a user-independent system.

Due to the last-mentioned aspect, this work aims to avoid the use of large training data sets and uses minimal data sets instead. As the term "minimal training data" is very vague, its concrete meaning in this context is explained in the following.

4.1.1 Meaning of "Minimal Training Data"

When talking about minimal training data, the following three main aspects have to be taken into account.

4.1.1.1 Amount of Training Data

In general, the amount of necessary training data heavily depends on the problem as well as on the approach used. Approaches can range from parametrized models based on process descriptions to representative data sets describing the problem observed. Consequently, the amount of training data can also range from simple, one-time measurements to large, statistically significant data sets.

For example, several activity recognition systems rely on training data sets containing multiple repetitions of each relevant activity. A real-life bicycle maintenance scenario introduced in [Ogr09] for example uses 20 repetitions for each of the gestures considered, which results in 3035 training instances and 291 minutes of training data. The same procedure was used in another experiment related to car assembly (shown in the same work). It is obvious, that the collection of such large amounts of training data is difficult to perform in real-life and large-scale scenarios. The data collection process for many computer vision approaches is even more complex. In some applications hundreds or thousands of training images were used in order to train recognition systems reliably (e.g. [Dal06] [SRE+05] [ZBMM06]).

However, approaches were also introduced which aim at reducing the amount of training data. In computer vision applications for example, so called "one-shot" approaches use quite restricted and minimal amounts of training images (e.g. [FFFP03] [FFFP06]). Other approaches as described in [Bla11] or [DBL11] show different ways to reduce the amount of training data.

Besides supervised learning methods as have been mentioned so far, semi-supervised learning (e.g [CSZ06]) was introduced to reduce the amount of labeled training data needed. Systems based on concepts such as self-training (e.g. [CSZ06]) and co-training (e.g. [BM98]) have been already used in many activity recognition scenarios (e.g. [SVLS08] [GYL+07]). Further examples and a more detailed discussion about this topic can be found in Section 1.1.2 and Section 1.2.2.

When talking about minimal training data sets in this context, the intention is to avoid large data sets and to use model-based approaches, "one-shot" training procedures and one-time parameter configurations as much as possible.

4.1.1.2 Necessary Effort to Collect Training Data

In addition to the amount of data being collected, the complexity of the data collection process must also be considered. Even if the system is configured by a few one-time measurements only, its usage in real-life applications is strongly restricted if such measurements can be exclusively performed by technical experts or with great effort. Consequently, in this context, the term minimal training data means, that simple, less time-consuming data collection procedures and measurements should be taken into account only. At best, it should be even possible for average citizens to perform the training data collection process themselves.

4.1.1.3 User/Environment Independent and Dependent Training Procedures

In general, one distinguishes between a user and environment-dependent and independent training process. The advantage of the latter approach is obvious: Systems can be pre-trained in labs or artificial environments for common usage. Consequently, they can be easily deployed in large-scale, real-world scenarios and can be used almost out-of-the-box even under various environmental conditions and by multiple users.

Simple recognition tasks like the detection of modes of locomotion can already be used out-of-the-box or based on simple one-time configurations (e.g. the Nike+ product portfolio[34] or the Jawbone[35] system). In contrast, many complex recognition systems rely on user- and/or environment-dependent training data sets in order to reach sufficient results. [Kun11] introduced four data sets ("House", "Opportunity", "Drink and Work" and "Bicycle"), on which a recognition system was trained in a user-independent way as well as for overall users. It is shown, that significant performance improvements were achieved in the case of "overall users". However, [OSLT08] introduced a data set (including 20 activities) and a system trained with user-dependent training data, that achieved similar results to a user-independently trained system applied on the same data set in [ZBS09].

Consequently, the success of user- and environment-independent training depends very much on the recognition problem as well as on the approach used. Due to the described advantages of such a system, the meaning of minimal training data in this context means the attempt to design systems that do not rely on user- and environment-dependent training sets.

In summary, the term "minimal training data" implies, that the systems introduced should fulfill the following requirements as much as possible:

- The amount of collected training samples should be reduced. The use of large amounts of data should be avoided as much as possible. In the best case, the systems should work out-of-the-box or rely on one-shot training procedures or simple threshold-based configurations only.
- Data collection should be performable with the least effort (simple collection/measurement procedures), in short time (e.g. one-time environment walk-throughs) and by technical laymen.
- The system should avoid user- and environment-dependent data sets as much as possible.

Figure 4.1 visualizes the idea of "minimal training data" in this context again.

4.1.2 Brief Overview

An on-body, core system (wrist-mounted camera with a proximity sensor mounted on top and on-body inertial sensors) is introduced in order to spot object interactions. It will be shown, that the system is able to outperform a state-of-the-art inertial system. Furthermore, the core on-body system is used as a starting point for several sensor fusion approaches focusing on a

[34] http://nikeplus.nike.com/plus/ (last accessed on 2013/05/10)
[35] https://jawbone.com/up (last accessed on 2013/09/02)

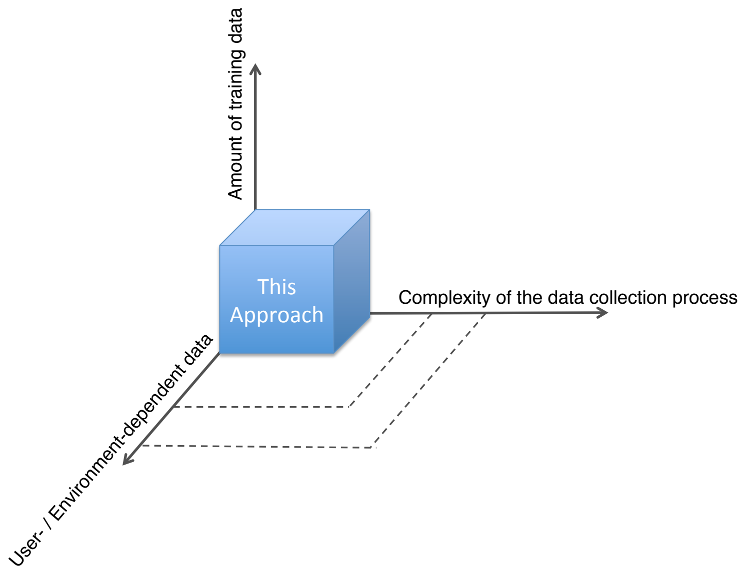


Figure 4.1: The meaning of minimal training data. This work tries to reduce the amount of training data, to perform only simple and less time consuming data collections and measurements as well as to use user-independent data.

search space reduction. Amongst other systems, ideas and concepts already introduced in Chapter 2 and Chapter 3 were considered. The applied systems meet the requirements of minimally intrusive instrumentation, minimal training data sets and low-cost sensor modalities. The introduced models are constructed from simple, one-time measurements performed by the person installing the system. An in-depth evaluation will show the impact of each fused system on the recognition performance. Finally, subtle hand activities are recognized on the basis of object interactions spotted in combination with motion features, smart appliances and the combination of both.

In the following, a detailed discussion about related work is given. Afterwards a state-of-the-art inertial approach is shown that was developed and applied by Ulf Blanke (ETH Zurich, Switzerland) on a data set that consists of subtle arm activities and related object interactions embedded in a large amount of day-to-day background activities. We will see, that the inertial system provides poor results in such a spotting scenario. As a next step, the contribution of this work and the general idea of the proposed system, applied processing and fusion algorithms are described. Finally, a concrete system implementation and in-depth system evaluations are shown.

4.2 Related Work

During the last few years a large number of methods focusing on the recognition of activities and object interactions has already been proposed. This section gives a detailed overview about state-of-the-art approaches related to this topic grouped by used sensor modalities.

4.2.1 Inertial Sensors

One of the most frequently used sensor modalities are acceleration sensors. Such systems focus on activities ranging from posture detection like "lying", "standing" or "sitting" (e.g. [FMT+99]

[JKC07]), over simple modes of locomotion such as "walking", "standing", "running", "going upstairs" and "going downstairs" (e.g. [RM00] [VLC00] [LM02] [MHS01]) to more complex daily life gestures such as "writing" or "shaking hands" (e.g. [Sti08] [AJT05] [ZS13]). Even sport gestures were focused on (e.g. [KBH+06] [HKG+06]). In [VLC00] acceleration sensors were attached to the outside of the upper leg (just above the knee). The objective was to recognize activities such as "running", "jumping", "climbing stairs" or "riding a bicycle". Both training as well as test data were recorded. Classification rates between 42% and 96% were achieved by a Kohonen map classifier. Activities like "climbing stairs" and "descending" were hardly recognized. On the contrary, activities such as "riding a bicycle", "standing" and "sitting" achieved recognition rates of about 90%. In [AJT05] the issue of detecting eating and drinking gestures is covered. Acceleration sensors were mounted on the upper and lower arm of participants. In a first step continuous data is segmented into motion segments. A sliding window and a bottom up (SWAB) approach were used to derive feasible gesture segments. In order to identify potential gestures, a Euclidean distance based similarity search was performed on segmented data. The final recognition of relevant gestures is done using HMMs. Experiments have shown that eating and drinking gestures could be distinguished from other movements with an accuracy of 95% (on isolated data). [KBH+06] deals with the topic of recognizing Tai Chi gestures using on-body motion sensors. All in all, eight acceleration sensors were attached to the human body (e.g. lower and upper arm, knee and neck). The objective was to distinguish between two common Tai Chi gestures ("repulse the monkey" and "parting the horses mane") as well as to distinguish between experts and amateurs. Using 10-fold cross validation and a kNN classifier, the proposed system achieved recognition rates of between 76% and 85%.

4.2.2 Wearable Cameras

Other techniques aim at the recognition of activities and object interactions on the basis of wearable cameras (e.g. [SSP98] [YOI92]) and eye tracking systems (e.g. [IMMR10] [TKSD12]). A common problem of using cameras to identify objects is, that computer vision algorithms need a large amount of training data that covers objects considered from different perspectives and different zoom levels. HOG features are widely used in such scenarios (e.g. [DT05] [LLZ+11] [ZZS07] [Dal06] [EZW+06]). In [DT05] the issue of human detection is focused on, whereas [LLZ+11] shows an evaluation based on the ImageNet dataset (see [DDS+09]), which includes 1000 object classes. Systems based on HOG were also introduced to handle the problem of recognizing texture-less objects (e.g. [HLI+10]). Besides, SIFT based techniques (e.g. [SRE+05] [PWF09] [Low99] [KHP07] [HHC+11] [HCI+12]) are also often used for object recognition applications as they are invariant to rotations, scale and illumination changes.

This thesis will introduce a wrist mounted camera as part of a multi-modal approach. HOG features are combined with a simple one-shot training (only one training image per object is used) to recognize object interaction. With the help of sensor fusion techniques, the impact of problems that usually occur in real-life applications (e.g. a very small training data set and changing lighting conditions) is minimized.

4.2.3 RF Systems

Widely used approaches focusing on the detection of object interactions and the recognition of human activities are based on radio-frequency identification (RFID) tags ([SHVLS08] [PFKP05] [CNL10] [WOC+07] [PFP+04]). In [CNL10] the iActionLogger system is introduced. The objective of the proposed system is to detect human-object interactions. 2.4 GHz radio modules were placed in rooms and attached to objects such as a table, printer, various coffee mugs and a coffee machine. Individual interaction radii were set for different devices and active nodes were carried in trouser pockets. One of the goals was to measure the object interaction time. It is shown that this task was solved with quite varying success as recognition rates ranged between 44% and 90%. In [PFP+04] the recognition of day-to-day activities on the basis of object interactions is considered. There, RFID tags were attached to common items such as a plastic fork and users had to wear a glove-based RFID reader. The idea was to use sensors in order to detect object interactions in a first step. Afterwards, a probabilistic engine was applied that infers activities based on sensor observations. Finally, a model creator was used to define probabilistic activity models. To evaluate the system, a real house scenario was considered in which 108 RFID tags were placed. 14 people were asked to carry out day-to-day activities. The objective was to recognize 14 daily life actions such as "oral hygiene", "washing", "telephone use" or "infant care". In this scenario the system achieved recall values between 33% and 93% and precision values between 64% and 100%. In general, all objects have to be instrumented with RFID tags and a RFID scanner is usually mounted on the user's hand or the body. The major disadvantages of such systems are that the reading range is often limited, which makes it hard to differentiate objects which are located close to each other and that the instrumentation effort is rather high.

4.2.4 Capacitive Sensors and Magnetic Field Measurements

Apart from common techniques such as cameras, RF-based systems or inertial sensors, [LARC10] [CAL10] [CBL12] propose capacitive sensing systems for various activity recognition applications. The observed scenarios range from the detection of complex hand gestures (identification of letters and numbers that users "write in the air") to "swallowing", "speaking" and "chewing" actions. Although the proposed results are very promising, no evaluations on subtle and barely distinguishable activities (e.g. pushing buttons or opening appliances) have been done so far. In [PSKL08] another interesting system, that is able to recognize Tai Chi gestures, is introduced. Using magnetometer resonance between sensors, the system can track the relative position of body parts as well as their orientation. First results are very promising as the system achieves better classification rates than a state-of-the-art inertial sensor system. However, a long term evaluation is missing so far. Besides, both techniques (capacitive sensing and magnetometer resonance) are based on a significant amount of training data, which makes it hard to use the system in large-scale scenarios where a large amount of various human activities can be of interest.

4.2.5 Smart Devices

Some approaches such as [LNV+06] [TIL04] use information coming from smart devices (such as binary switches or mainstream devices equipped with proximity sensors) to recognize object interactions and underlying human activities. A big disadvantage of such approaches is, that almost all mainstream objects are unable to deliver their current operating modes in a standardized way. Consequently, additional infrastructure equipment is needed. However, the most important disadvantage of such systems is, that they are not able to identify the person using the device.

4.2.6 Sound Analysis

Systems based on sound analysis are also often used to recognize human activities. In [AKT07a] [AKT07b] [CKZ+05] [FLKB09] systems are introduced that analyze environmental sounds. In [FLKB09] sounds coming from nine office and household appliances such as printers, copiers or coffee machines were recorded using an iPhone. The mobile phone was placed in several on-body locations such as in a trouser pocket, jacket pocket or on a belt. The objective of this work was to compare recognition rates from individual locations. Sound data was processed using FFT, LDA and a kNN classifier. To evaluate the system 15 recordings were performed whereas five recordings were used to train the classifier. The system achieved very good recognition rates of above 90% when it was trained and tested at the same location. However, when training the system with clean data (the iPhone was held in the hand) the classification rate decreases to around 60%. In [CKZ+05] an automatic bathroom activity monitoring system based on sound analysis is introduced. The objective is to generate reports on personal hygiene including five activities such as "showering", "brushing teeth" and "urinating". Therefore, a small omni-directional microphone was installed close to the washbasin. All in all 132 sound samples were recorded by five people during several experiments. Based on the recorded sound data Mel-Frequency Cepstral Coefficient (MFCC) features were calculated and applied to an HMM. In a simplified scenario, recognition rates of between 87% and 93% were achieved.

4.2.7 Multi-Modal Approaches

A large number of systems are using multi-modal sensor approaches (see [NCLZ11] [MKYS12] [MYK+10] [WNS06] [WLTS06a] [OSLT08] [SHVLS08] [WPP+07] [BPPW09] [SLT07] [RFC+09] [Bla11]). In [NCLZ11] and [MKYS12] similar approaches to this work are shown. In [NCLZ11] a multi-modal system, which is based on a wearable camera, acceleration sensors and a microphone (the last two sensors were utilized via a Nokia N95 mobile phone), is introduced. The objective is to show that the proposed approach is able to provide similar recognition rates when using only 9% labeled data as state-of-the-art systems that are using 100% unimodal labeled data. The system consists of three phases. First, a collaborative data collection is performed where a small amount of labeled data and a large amount of unlabeled data is collected. Based on raw data features such as mean, standard deviation (in terms of accelerometer sensor), a number of pixels of representative colors (in terms of image features) and a 13 order MFCC (in terms of audio processing) are calculated. In the second phase classifiers are trained with labeled data from the three sensors. After that, they cooperate with each other based on unlabeled data and enhanced co-training algorithms. Finally, they are refined iteratively. In the last phase the classifier's outputs are combined to improve the system's accuracy. The system was evaluated in a real-world environment covering 13 activities such as "taking the bus", "going up/down in a lift" or "sitting down". Besides, it was compared to well known semi-supervised state-of-the-art approaches. In [MKYS12] the multi-modal WristSens system is shown. The system worn on the wrist includes a wearable camera, an accelerometer and a microphone. The camera is used to detect objects that are located in the user's hand. The idea is that used objects may relate to specific activities. Hand motion data is analyzed in order to recognize performed gestures. Finally, the microphone is used to recognize sounds that are related to specific activities. Hence, the system combines motion features (mean, energy, frequency-domain entropy and dominant frequency), image features (color histograms) as well as sound features (Mel-Frequency Cepstral Coefficient). The recognition is performed by using decision tree classifiers. The objective was to detect daily life activities such as "making coffee" and "brushing teeth" in real-time. Unfortunately, a detailed system evaluation is missing. In [MYK+10] the authors extended their work shown in [MKYS12]. In addition to the sensors already proposed, they integrated an illuminometer as well as a digital compass. Raw data from these sensor modalities were used directly as features. The objective was to recognize 15 daily life activities such as "watering plants", "making juice" or "listening to music". To evaluate the system, data was recorded from 10 users in two experimental environments. The evaluation process uses a leave-one-session-out cross validation and considers AdaBoost M1 and a C4.5 decision tree in combination with HMMs (instance based and window based). The best results for both experiments were achieved by C4.5+HMM (instance based; recall and precision values of more than 84%). Afterwards, a more detailed evaluation on the impact of each sensor modality is shown. The biggest impact was achieved by the camera sensor, followed by the accelerometer and the microphone. The illuminometer as well as the compass barely contributed to the recognition quality. Compared to [NCLZ11] [MKYS12] [MYK+10], this thesis uses different sensor modalities. Furthermore it evaluates the impact of the achieved search space reduction through sensor fusion approaches on the recognition quality, as well as on the system's performance in much greater detail. Moreover, the thesis is exclusively based on low-cost approaches which were designed to work with a minimal training data set. This fact makes the system usable in large scale and real-world applications.

4.3 State-Of-The-Art Inertial Systems and the Problem of Spotting Subtle Hand Activities and Related Object Interactions

As was already discussed in Section 4.2, inertial sensor based systems are a widely used solution to spot hand gestures and to recognize human activities (e.g. [RM00] [VLC00] [MHS01] [Sti08] [ZS13] [KBH+06]). Although inertial sensors are able to provide excellent recognition rates for rather characteristic activities, they may fail when it comes to subtle and barely distinguishable hand activities which are the focus of this work.

Figure 4.2 and Figure 4.3 affirm this first assumption. Figure 4.2 shows raw acceleration signals (generated by a wrist mounted 3-D motion sensor) of two activities that were recorded within the Skoda data set (see [OSLT08]). It can be seen that even raw motion signals are quite similar for performed activities of the same class and are quite different among different classes. Consequently, it should be feasible to achieve good classification results by analyzing motion data. This assumption was confirmed by many publications that have focused on this data set (see [OSLT08] [ZWS09] [MBP14] [SRO+08]).

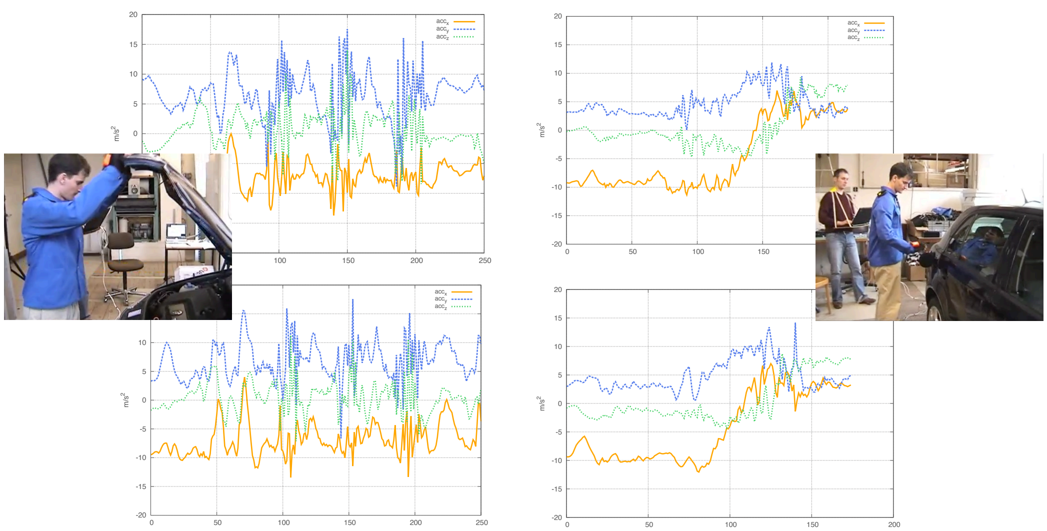


Figure 4.2: Acceleration signals generated by a wrist-mounted motion sensor. Raw signals and images are taken from the Skoda data set described in [OSLT08]. Left: Two repetitions of the activity "Open hood"; Right: Two repetitions of the activity "Open gas cap".

In contrast, Figure 4.3 shows signals generated by the subtle activities "Press light button" and "Press printer button". Although signals from different classes vary, the difference is less pronounced and not as obvious as in the previous example. Moreover, the signal variance within the same class is quite high and the variance between different classes is not very distinctive. Consequently, it will be much more challenging to reach good classification results in the case of subtle, barely distinguishable activities.

In the following, it is shown that an inertial sensor based on a state-of-the-art approach provides poor results for such a spotting scenario although it is using a statistically significant, large training data set. The inertial system used was developed and applied by Ulf Blanke (ETH Zurich, Switzerland) and is evaluated on an office data set that contains several subtle arm activities and related object interactions which are embedded in a continuous stream of background data. In contrast, this work will introduce a system that is able to outperform this approach significantly and is therefore able to deal with such activities. Moreover, the approach shown is based on simple one-time configuration and measurements only.

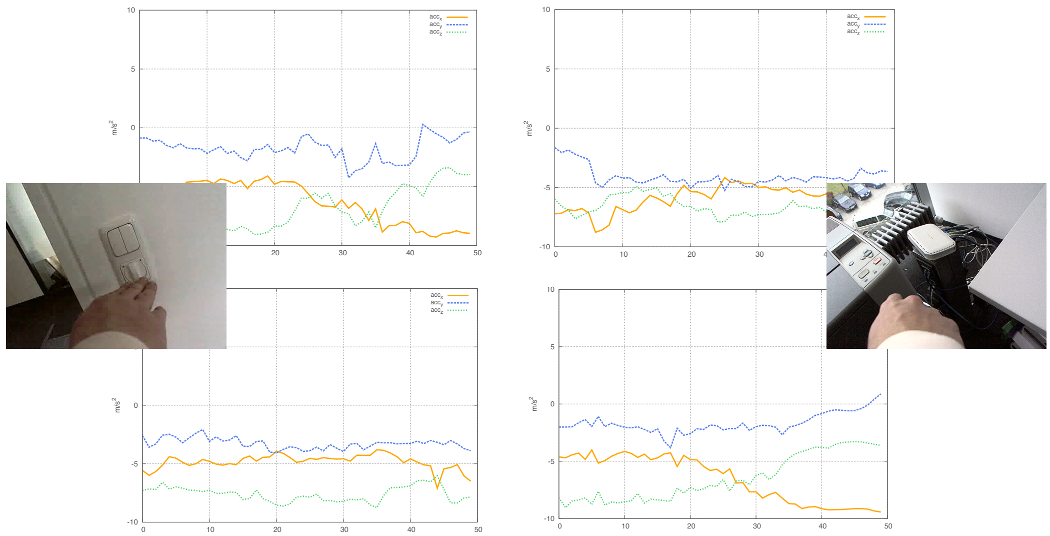


Figure 4.3: Acceleration signals generated by a wrist-mounted motion sensor. Left: Two repetitions of the activity "Press light button"; Right: Two repetitions of the activity "Press printer button".

4.3.1 Inertial Sensor Approach

The system uses three inertial sensors attached to the participant's body (see Figure 4.4). Two sensors were attached to the person's arm – one on the forearm and one on the upper arm. In addition, a sensor was attached to the person's back. Each sensor provides a twelve dimensional feature vector containing information about acceleration (three axes, unit: m/s^2), angular velocity (three axes, unit: deg/s^2) and the magnetic field (three axes, unit: $mGauss$). Furthermore, each sensor delivers information about its orientation (Euler angle, three axes, unit: deg). The inertial system is hereafter called IS.

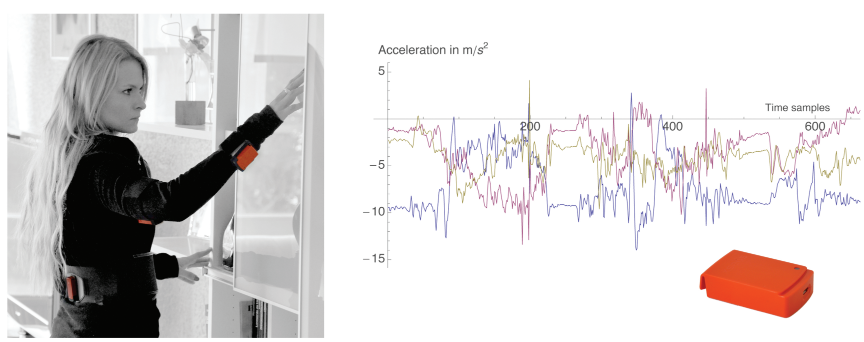


Figure 4.4: Left: Xsens acceleration sensors fixed to the human body: Forearm, upper arm and back (color coded). Right: Xsens sensor and corresponding 3-dimensional acceleration signal (forearm mounted sensor) generated through hand movements.

4.3.1.1 Segmentation and Classification Methods

The segmentation procedure, feature calculations as well as a frame based classification were performed by Ulf Blanke (Wearable Computing Laboratory, ETH Zurich). The following three paragraphs summarize his work which can also be found in [BBLS13].

Common steps were performed (as described below) to segment the continuous training data stream into regions. Based on these regions mainstream features were calculated and fed together with the corresponding labels into a learning process of a classification model. During the classification process, the continuous data stream was segmented in candidates feasible for containing an activity. After that, features were extracted and using the trained classification model scores for all classes (activities) were obtained.

Basic Segmentation Procedure Several segmentation approaches exist to group a continuous data stream into segments for an activity (see [ZVLS07] [LX96] [AJT05]). The most common approach is based on a sliding window technique with a fixed window length. This approach was also applied in this work. The window length was estimated from the mean activity duration. In general, a fixed window size might not be the best choice as it can influence the recognition quality (see [HS05]). Nevertheless, this approach was chosen as baseline since it worked well for a variety of activity recognition systems (see [BS09] [HFS08] [BI04] [DD01] [VLC00]). In the following, a fixed window length of 60 samples and a sliding window technique with an overlay of 5 samples was used.

Feature Calculation For each calculated segment widely used features (e.g. mean, variance and frequency space based features) were extracted (see [LCK+05] [BI04] [BSK+10]). The feature space of the training data set was standardized and the parameters obtained were used for the test set. Each feature was calculated independently for each dimension of acceleration, gyroscope and orientation data. The following list shows the feature types in question:

- Coefficients – grouped in 5 exponential bands
- Cepstral coefficients
- Spectral entropy
- Mean and variance
- Cumulative energy

Classifier As classifier paradigm support vector machines with a radial base function kernel were used (see [CL11]). To overcome the multi-class problem, a one-vs-one learning procedure was used. Regularization parameters were obtained experimentally from a training data set. The training data set consists of 15 repetitions for each activity which were performed by a single person who did not take part in the experiments. This way the system became user independent.

During the learning phase random segments were extracted from the background class. There, the number of negative samples was equal to the number of positive samples of all classes together. After that the classifier model was trained to deliver probability estimates between 0 and 1 for each class. Each classification step delivers a normalized score vector per segment containing calculated scores for the activities focused on during this work. In order to handle multiple overlapping windows that were created by the sliding window segmentation procedure, a non-maximum suppression technique was used. For each timeframe, overlapping windows were selected and based on them the maximum score for each activity was calculated. This value was used as final activity score for the corresponding timeframe. Finally, the activity label with the highest score was obtained for each timeframe.

Final Segmentation Procedure The obtained scores per timeframe were used to calculate new segments for each class. Therefore, segments with a fixed minimum length of more than 1/3 of a second and a fixed maximum length which was less than 20 seconds were defined. Within each segment all scores had to exceed a SVM score threshold ($IS_{thrSVMScore}$). The minimum and maximum length was chosen with the help of the standard durations of all the activities in focus. Afterwards, nearby segments were fused. The time difference between two consecutive segments was used as a distance measure. During an optimization process, several time thresholds $IS_{thrFusion}$ ranging from 0 to 1 second in steps of 0.2 seconds were evaluated (the optimized inertial sensor system is referenced as IS_{opt} in the following).

4.3.2 Data Set: Subtle Hand Activities and Object Interactions within an Office Scenario

As a next step, a data set is introduced consisting of subtle arm activities and related object interactions that were recorded within an office scenario. Activities of interest were embedded in a continuous stream of daily-life background data. The data set is used as an example showing that the usage of inertial systems in such spotting scenarios is quite limited. In the following, a detailed description of experimental conditions, participating people as well as the environment in which all experiments were carried out is given. Besides, the considered activities of interest, resulting object interactions and the performed background activities are introduced.

4.3.2.1 Environment Description

An office environment was chosen in which all intended experiments were carried out under conditions similar to real life. The area consists of thirteen rooms including ten office spaces, a kitchen, a printer room and a meeting room. Within that space employees were allowed to move around freely. Furthermore, they were told to follow their normal working routine. In contrast to this, experiment participants were only allowed to move within the observed area. This region included the hallway as well as an office space, the kitchen, the meeting room and the printer room. Figure 4.5 shows the department's floor plan and the observed area.

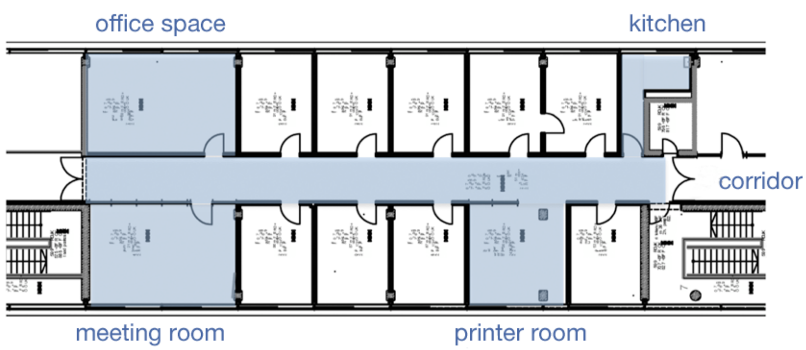


Figure 4.5: Office department: The monitored region is marked in blue.

Six students (four male, two female) were invited to join the experiments. Each of them had to repeat a list of actions including both activities of interest and so-called background activities several times. Between carrying out activities and while walking around, participants were allowed to move as they wanted to. This way the pre-defined activity list was repeated 27 times (in the following repetitions are called runs). The activity sequence for each run was randomly generated while observing the logical flow with respect to dependent activities and rooms.

4.3.2.2 Subtle Activities and Object Interactions

The objective was to spot and recognize 32 different hand activities within a continuous data stream including daily life activities (The ratio between activity of interest and background activity samples is 1:10). The recognition problem is made more complicated by the fact that the chosen activities are subtle and barely distinguishable hand actions and by considering almost similar hand activities used for different purposes.

Apart from the recognition of subtle hand activities the goal of this work is to also spot and recognize the resulting object interactions. For many activity recognition problems such high-level information is already sufficient. For example, the fact that a microwave was used is already enough information to learn user behavior patterns and consequently the recognition of specific underlying hand activities (e.g. use-mode detection) is not needed. While performing activities of interest, the participants were in touch with 16 different objects. Table 4.1 lists the activities of interest and their number of repetitions grouped by underlying objects and their location[36]. All in all 713 object interactions based on 16 different object types were performed.

Table 4.1: Performed object interactions and activities of interest grouped by rooms: kitchen, printer room, meeting room and office

Object	Activities (Repetitions)
Microwave	Open (27), Close (27), Start (11), Clean (16)
Coffee Machine	Make Espresso (14), Make Coffee (13)
Power Socket	Connect Cable (27)
Cupboard	Open (27), Close (27)
Wall Cupboard	Open (27), Close (26)
Ethernet Connector	Connect Cable (30)
Water Tap	Fill Big Cup (14), Fill Small Cup (13)
Battery Charger	Put In Empty Battery (15)
	Remove Battery While Charging (14)
Laser Printer	Take Printout (12), Push Button (15)
Ink Printer	Take Printout (15), Push Button (12)
Climatic Control	Turn Left (13), Turn Right (14)
PC	Turn On (55)
Scanner	On (29), Off (27), Scan Document (27)
Air Conditioner	On (27), Off (27)
Light-Shutter Switch	Light Button - Switch (28),
	Shutter - Turn Left (8), Shutter - Turn Right (19)
Ring Binder	Take Binder Out (28), Put Binder Back (28)

Besides activities of interest, each participant had to perform a set of background activities. This set includes daily life activities which are usually performed in an office department such as cleaning a whiteboard, opening a window or stirring coffee. Table 4.2 lists background activities and the resulting object interactions grouped by their location. Each background activity was carried out once per run.

In addition, participants were allowed to perform any kind of event-related day-to-day activity during the experiment. This means that some activities were performed because it was necessary due to the current situation (e.g. refilling the coffee compartment or opening a milk bottle). Such unpredictable activities are not listed in the above table.

All in all almost 7 hours of data was recorded. About 91% of the time can be accounted for background activities, which highlights the difficulty and importance of the spotting problem once more.

4.3.3 Evaluation

In the following, the evaluation of the inertial sensor system for both scenarios (the spotting of object interactions and underlying, subtle arm activities) is shown. Before the achieved results are shown, the overall evaluation procedure is introduced.

[36]Note: The microwave was cleaned inside. Consequently, each cleaning event was performed immediately after a "microwave opened" event.

Table 4.2: Background activities and resulting object interactions grouped by rooms: kitchen, printer room, meeting room and office

Object	Activities
Cup	Drink from cup, stir up coffee
Milk Bottle	Take milk from fridge
Newspaper	Read newspaper
Table	Clean table
Printout	Put printout on wall
Whiteboard	Clean whiteboard, write on whiteboard
Door	Open, close
PC	Log in and print document
Coins	Count
Table	Move items on table
Window	Open, close
Map	Point finger on wall mounted map
Picture	Admire picture on wall

4.3.3.1 Evaluation Procedure

A common evaluation procedure based on precision, recall and equal error rate (EER) was used to evaluate the system. The following values were investigated:

- Overall System Evaluation: The highest reachable average recall ($recall$), the corresponding average precision ($precision$) – both based on results achieved for objects/activities – and the EER (EER) were calculated. Besides, average values of $recall$, $precision$ and EER were considered.

- Object/Activity Related Evaluation: Recall ($recall_{obj}$), precision ($precision_{obj}$) and EER (EER_{obj}) values were calculated for each object/activity. Besides, average values of $recall_{obj}$, $precision_{obj}$ and EER_{obj} were considered.

To calculate recall and precision, true positives were defined as an overlap between the recognized objects/activities and the corresponding labels. Every recognized object/activity without a corresponding label was defined as a false positive. The EER value was defined by the point where recall and precision have exactly the same value. As this point might not always exist, a linear interpolation between the existing recall-precision values was used to calculate the EER values. All evaluations were performed "offline".

4.3.3.2 Evaluation: Spotting Object Interactions

In a first step, the inertial system was used to spot object interactions. Therefore, the system was trained based on the introduced training data set and using object interaction labels. Hence, 16 object classes were considered.

Furthermore the impact of $IS_{thrSVMScore}$ and $IS_{thrFusion}$ on the overall recognition quality was evaluated. Several threshold values for $IS_{thrSVMScore}$ ranging from 0 to 1 in steps of 0.0015 as well as for $IS_{thrFusion}$ ranging from 0 to 1 second in steps of 0.2 seconds were analyzed. Figure 4.6 shows the resulting $(1 - precision) - recall$ curves.

The highest $recall$ of 80% could be achieved with $IS_{thrSVMScore} = 0.0135$ and $IS_{thrFusion} = 0$ seconds (i.e. no segment fusion). The corresponding $precision$ of 3% is very low making the system unsuitable for almost all application areas. This fact is confirmed by a rather low EER of 18%. The reason for this is clear: Characteristic motion patterns which are the most important precondition for inertial based recognition systems are missing.

However, when analyzing the recognition quality for specific objects, it can be seen that for two objects ("Climatic Control Panel" and "Wall Cupboard") reasonable equal error rates

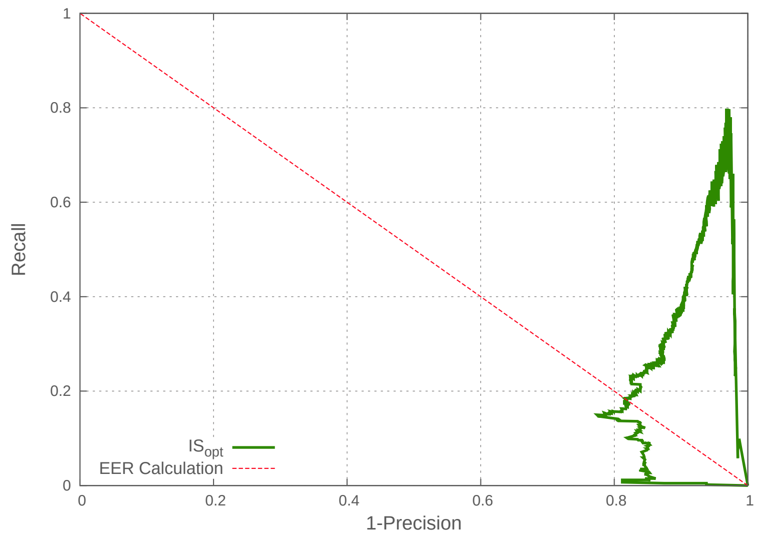


Figure 4.6: $(1 - precision) - recall$ curve: IS_{opt}

of 70% and more can be achieved (see Table 4.3). This means that these activities include characteristic motion patterns strong enough to separate them well from the background and all other activities.

Table 4.3: Object overview for IS_{opt}: $recall_{obj}$, $precision_{obj}$ and EER_{obj}.

Object	$recall_{obj}$	$precision_{obj}$	EER_{obj}
Battery Charger	86	2	0
Coffee Machine	74	2	0
PC	87	3	0
Air Conditioner	32	4	0
Climatic Control Panel	100	2	74
Microwave	60	4	38
Ethernet Connector	97	2	0
Ring Binder	95	8	21
Power Socket	78	1	11
Laser Printer	85	1	4
Ink Printer	93	2	11
Light-Shutter Switch	73	3	14
Scanner	68	3	1
Wall Cupboard	100	6	70
Cupboard	80	4	26
Water Tap	70	2	0
ØAverage	80	3	17

4.3.3.3 Evaluation: Spotting Subtle Hand Activities

It was shown that the inertial system achieved poor results when spotting object interactions. Consequently, similar or even worse results are expected if the underlying subtle hand activities are to be spotted. For this scenario the system was trained based on the same training data set, but using activity labels instead of object interaction labels. Some activities were joined together due to the fact that they can't be distinguished based on motion patterns only (e.g. turning a device on or off). Table 4.4 lists the merged activity set including 27 different subtle hand activities.

Table 4.4: Merged activity set I: 27 activities grouped by locations (kitchen, printer room, meeting room and office)

Object	Activities (Repetitions)
Microwave	Open (27), Close (27), Start (11), Clean (16)
Coffee Machine	Make Coffee (27)
Power Socket	Connect Cable (27)
Cupboard	Open (27), Close (27)
Wall Cupboard	Open (27), Close (26)
Ethernet Connector	Connect Cable (30)
Water Tap	Fill In Cup (27)
Battery Charger	Put Empty Battery (15)
	Remove Battery While Charging (14)
Laser Printer	Take Printout (12), Push Button (15)
Ink Printer	Take Printout (15), Push Button (12)
Climatic Control	Turn Left (13), Turn Right (14)
PC	Turn On (55)
Scanner	Use Scanner (83)
Air Conditioner	Use Device (54)
Light Switch	Push Light Button (28)
Shutter Switch	Shutter - Turn Left (8), Shutter - Turn Right (19)
Ring Binder	Take-Put Binder Back (56)

Figure 4.7 shows the results achieved. As expected the inertial sensor system is not able to provide a reasonable recognition rate. The highest *recall* of 84% could be achieved with $IS_{thrSVMScore} = 0.009$ (evaluated from 0 to 1 in steps of 0.0015) and $IS_{thrFusion} = 0$ seconds (evaluated from 0 to 25 seconds in steps of 1 second), which means that no segment fusions were performed. The corresponding *precision* is only 2% and the *EER* value is 13%, which is 5% lower than was achieved by the system recognizing object interactions. Nevertheless, the achieved maximum *recall* could be raised by 4% to 84% while loosing 1% *precision*. Table 4.5 shows the results for specific activities confirming the fact that an inertial system achieves poor results for the kind of spotting problem observed.

4.3.4 Conclusion

We have seen that a state-of-the-art inertial system achieved poor results (an EER of 18% and less) in the case of spotting subtle, barely distinguishable hand activities and related object interactions. The scenario of spotting object interactions could be solved with a maximum *recall* of 80% and a corresponding *precision* of only 3%. The achieved *EER* was 18%. It was shown that only two out of 16 objects were spotted with an acceptable EER_{obj} of more than 70% ("Climatic Control Panel", "Wall Cupboard"). Even worse results were achieved in the case of spotting subtle hand activities. There, the highest *recall* was 84% with a corresponding *precision* of only 2%. The *EER* was 13%. Again, only two activities were spotted with an EER_{act} of more than 70% ("Shutter:Close", "Wall Cupboard: Opened").

To sum up, the inertial system is able to reach acceptable *recall* values of 80% and more in both scenarios. However, corresponding *precision* values were rather low (below 5%).

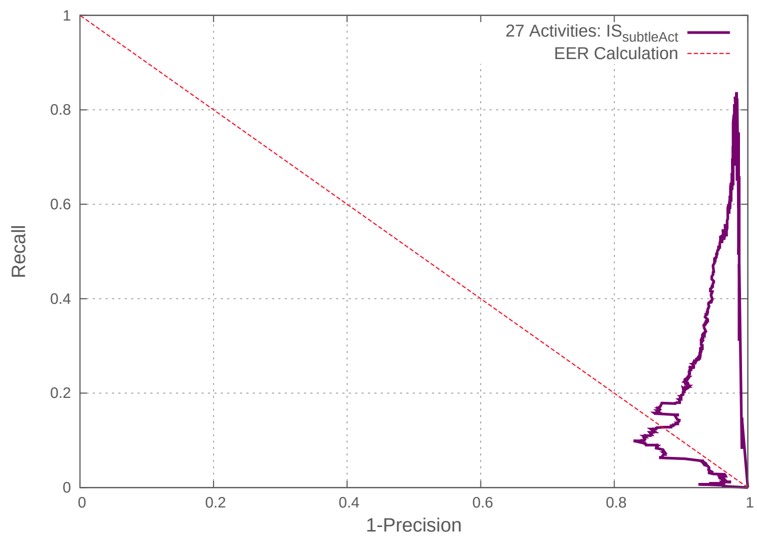


Figure 4.7: $IS_{subtleAct}$ (purple curve; inertial system) : $(1 - precision) - recall$ curve.

Table 4.5: Activity overview for $IS_{subtleAct}$: $recall_{act}$, $precision_{act}$ and EER_{act}.

Activity	$recall_{act}$	$precision_{act}$	EER_{act}
Battery Charger: Put Battery In	87	1	7
Battery Charger: Remove Battery	79	1	0
Coffee Machine	82	2	0
PC	80	3	0
Air Conditioner	67	4	0
Climatic Control: Turn Left	100	1	15
Climatic Control: Turn Right	100	1	57
Shutter: Open	75	1	0
Shutter: Close	85	1	70
Use Light Button	93	2	14
Microwave: Clean Device	88	1	0
Microwave: Device Closed	96	2	12
Microwave: Start Program	55	1	0
Microwave: Device Opened	67	1	15
Ethernet Connector	80	2	0
Ring Binder	100	5	21
Power Socket	70	1	11
Laser Printer: Take Printout	75	1	0
Laser Printer: Push Button	80	1	0
Ink Printer: Take Printout	100	1	7
Ink Printer: Push Button	92	1	0
Scanner	55	3	1
Wall Cupboard: Closed	96	3	0
Wall Cupboard: Opened	100	2	74
Cupboard: Closed	82	2	0
Cupboard: Opened	85	2	33
Water Tap	93	2	0
ØAverage	84	2	12

4.4 Research Questions and Contribution

The previous section indicated that the exclusive use of on-body motion sensors does not provide sufficient information to reliably spot the kind of activities targeted by this thesis. Although recall values of 80% and more were achieved, the corresponding precision values were very low (below 5%). Besides, the equal error rates (EER) were 18% and less. This fact shows that motions related to subtle hand activities are quite similar to motions regularly performed during other activities. Consequently, they can't be recognized based on noisy signals of on-body motion sensors and a limited amount of training data. As a consequence, the ***first research question*** that is addressed in this work is:

What other unobtrusive wearable sensor modalities can be used to solve the kind of spotting problem considered? Which algorithms can be applied on related sensor data without the need for large amounts of statistically significant training data?

As is shown in Section 4.7.1, sensor modalities and algorithms, that could significantly improve the spotting performance, were identified. The core on-body system consists of a wrist camera with a proximity sensor mounted on top and on-body inertial sensors. However, the achieved EER (47% in case of object interactions) of this basic fusion approach is still too low for almost all applications. Therefore, the ***second research question***, that was addressed in this work is:

Can the achieved results be improved by fusing the sensors and algorithms mentioned above with concepts and systems introduced in Chapter 2 (region of interest location) and Chapter 3 (recognition of device use-modes)? What are the most appropriate fusion techniques that do not require large amounts of training data?

Although object interactions and corresponding subtle hand activities are related to a device's operating mode and/or location, they are definitely not fully determined by them. With respect to subtle hand activities causing operating mode changes, the reasons are:

- Not all of the considered interactions are based on devices that can be unobtrusively turned into smart appliances able to provide operating mode changes (e.g. a ring binder or a cupboard).
- The fact, that this work focuses on a multi-user environment in which several people are interacting with the same devices. Consequently, a devices' operating mode change does not imply that the monitored user was interacting with this device.

With respect to sub-room level locations, the reasons are:

- Two-dimensional regions of interest may include multiple objects, that are located close to each other. Besides, they also cover the area around specific objects and object groups. Consequently, the fact that somebody is located within such a region of interest clearly does not infer that he/she intends to perform a specific object interaction.
- This work considers the idea of recognizing activities within regions of interest, which results in an anonymous, sub-room level based location system. Due to the fact, that a multi-user environment was in focus (several people are interacting with objects of interest), the information that somebody is currently located within an area around a specific object or a group of objects does does not imply that the monitored user is that person.

Apart from sub-room level location systems, several state-of-the-art approaches exist to determine the position of people indoors. Because the accurate three-dimensional location of the user's hand is usually determined by expensive and obtrusive systems based on intense calibrations (e.g. Ubisense[37], Lukotronic[38] or "Flock Of Birds"[39] system), they were not considered in this work. However, the impact of a less accurate, but easy to deploy room-level location system was evaluated. Section 4.7.2 and Section 4.7.5 show that location systems as well as the recognition of device operating mode changes can significantly improve the spotting performance. However, due to the issues mentioned above they still do not lead to satisfying recognition rates (EER up to 60%). Consequently, the ***third research question*** addressed in this chapter is:

What other pre-conditions, that apply to subtle hand activities, can be measured with low training effort? How and to what degree can these pre-conditions be used to improve the system introduced?

This research question is motivated by the fact, that threshold-based methods or systems trained by less time-consuming, one-time environment walk-throughs can be used to describe generally valid activity preconditions. In many cases, such systems can nowadays be: (a) used out-of-the-box, (b) pre-trained in labs in a user and environment independent way or (c) easily be set up by people with non technical background. Moreover, systems are able to restrict the search space of the considered problem and consequently a performance improvement may be achieved while complying with requirements on minimal training data and easy system deployment. In detail, the idea was to evaluate the impact of fusing the system introduced with information about the user's motion, hand motion intensities, time features as well as magnetic field signatures.

As was noted in Section 4.3.4, an inertial approach was not able to provide sufficient precision values for the considered problem by itself although it is using large amounts of training data. However, such a system may contribute to an improved performance of the introduced system. This idea leads to the ***fourth research question***:

How and to what degree can a state-of-the-art activity recognition system based on statistical learning techniques improve the introduced system?

At this point the restriction on the use of large, statistically significant training data sets is temporarily softened. The objective was, to evaluate to what degree the usage of such data sets can improve the results achieved. Section 4.7.7 will show, that the inertial system is not able to contribute to the performance of the introduced system at all.

So far, the core on-body system was able to significantly outperform the state-of-the-art motion system. One of the systems' key components is a wearable, wrist-mounted camera used to identify objects on images based on pre-selected segments. The fact that image processing tasks are known to be computationally extensive and the precision of the inertial system may be increased if it is applied on a reduced search space, leads to the ***fifth research question***:

What is the impact of replacing the core component of the introduced system (wearable camera and one-shot training) with a motion system using a large training data set?

Two concepts were considered in Section 4.7.9. In a first step, only a minimal sensor setup should be used. Therefore the camera and the proximity sensor of the core on-body system were replaced by the inertial system. Consequently, information about the user's hand height (derived from a simple body model) was combined with motion information. In a second step, the idea was to use exactly the same spotting procedure as the core on-body system but this time using an inertial approach to identify object interactions instead of a wearable camera. It will be shown, that the latter approach achieved the best results and was able to raise the EER

[37] www.ubisense.net (last accessed on 2013/06/18)

[38] www.lukotronic.com (last accessed on 2013/06/18)

[39] http://www.ascension-tech.com/realtime/rtflockofbirds.php (last accessed on 2013/08/31)

of the state-of-the-art inertial system (shown in 4.3.1) by 9% to 27%. However, despite using large training data sets, the proposed system is still unable to compete against the core on-body system introduced. Consequently, the restriction of avoiding large amounts of training data is put in place again.

The ***sixth research question***, that was considered in this work, is:

What is the impact of fusing multiple systems which rely on minimal training data on the performance of the system?

This step was motivated by the fact, that fusing the core on-body system with single methods resulted in a significant performance improvement. Consequently, several combinations of the proposed system extensions were evaluated. For example, combining information of a region of interest-based location system and a room level positioning system, results in a significant search space reduction and may consequently lead to a significant performance improvement. Besides, the impact of each additional system and the "price that has to be paid" for using it (e.g. environmental instrumentation or additional sensor modalities) is shown.

So far, fusion approaches were introduced, that are able to provide sufficient results in the case of spotting and recognizing object interactions. However, the information about an object interaction event (e.g. air conditioner was used) may not be detailed enough for many applications like the recognition of user behavior patterns (e.g. user turned air conditioner on/off or increased/decreased the cooling level). This fact leads to the ***seventh research question***:

How and to what degree can subtle hand activities be recognized by the introduced system?

This step was motivated by the fact, that object interactions have already been identified. The key idea was, that information about such spotted events can be re-used and combined with the approaches already introduced in order to recognize the concrete underlying subtle hand activity. Therefore, concepts related to a model-based device use-mode recognition system and a state-of-the-art inertial system were considered.

4.5 Approach and Overview

This section describes the general idea behind the proposed system. In the following, the kind of considered activities/object interactions are investigated, that can be determined on an abstract level by considering the research questions introduced in Section 4.4:

- Firstly, every activity execution relies on a set of pre-conditions. The detection and usage of such pre-conditions is the basic idea of the proposed core on-body system and several related fusion approaches. It is related to the ***first*** and ***third research question***.

- Secondly, people perform concrete actions during an object interaction. It is obvious, that the actions in question are mainly related to hand gestures/activities. Thus, the impact of fusing a state-of-the-art motion sensor based activity recognition approach with the core on-body system was evaluated in the ***fourth*** and ***fifth research question***.

- Thirdly, every interaction process has an effect on the object. In case of appliances such effects often result in an operating mode change. Chapter 3 already showed systems that are able to recognize operating modes of common household appliances. These concepts and ideas are included in the ***second research question***. However, the effect on simpler objects (e.g. ring binders or cabinet doors) is mainly motion or displacement. In order to detect such effects, additional instrumentation is needed. One example is the opportunity dataset (see [LPB+10]) where forks, bottles and plates were equipped with motion sensors in order to detect such interaction events. As an unobtrusive environmental integration was one of the main requirements of this work, such systems were not considered.

- Fourthly, the combined consideration of individual characteristics/pre-conditions lead to a more accurate description of the performed activity/object interaction event. Therefore, various fusion approaches of the core on-body system with several of the proposed concepts are focused on in the ***sixth research question***. Besides, concepts considered in the ***second*** and ***fourth research question*** were used again to recognize concrete subtle hand activities based on already spotted object interactions in the ***seventh research question***.

In the following, concrete ideas regarding the research questions introduced are discussed in detail. Before that, Figure 4.8 gives a brief overview of considered concepts with respect to the spotting and identification of object interactions.

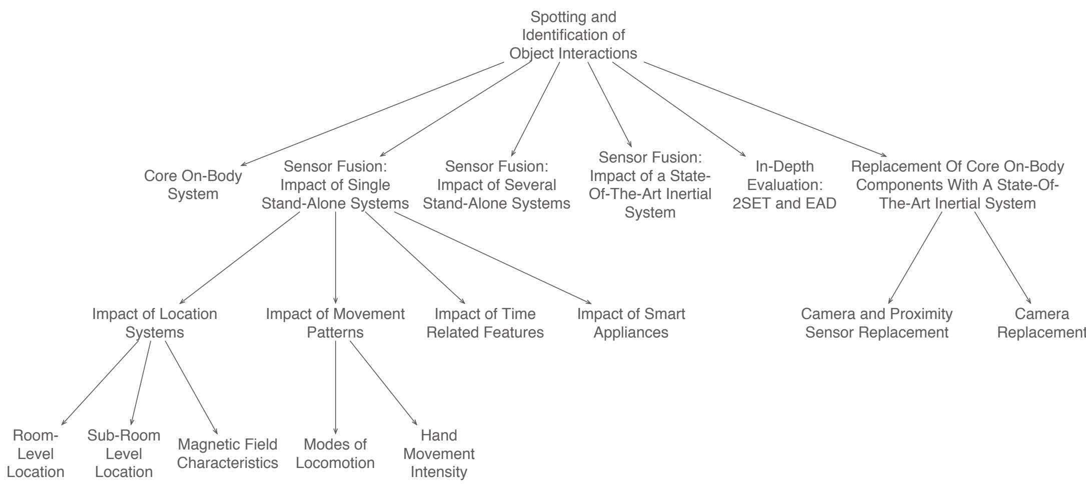


Figure 4.8: Overview: Spotting and identification of object interactions.

4.5.1 First Research Question: Wearable Sensing Modalities and the Core On-Body System

This work focuses on activities involving interactions between the user's hand and objects. Consequently, the key pre-condition is the proximity between the hand and an object. Several approaches can be used to detect such proximity based on wearable sensors. The three main concepts are:

- Arm worn RFID readers: In combination with RFID tags applied on relevant objects, such systems can detect whether or not the user is close to an object (see [CNL10] [PFP+04]). This concept has been widely used and its advantages and disadvantages have already been evaluated in depth. However, such an approach was not considered in this work due to the fact that relevant objects must be equipped with tags, which circumvents the idea of unobtrusive and minimal instrumentations. Besides, such systems have well-known problems related to the reading range of wearable RFID readers.

- Arm and object tracking: Object interactions can be inferred from location information of both the user's hand and objects. However, for such scenarios high precision tracking systems are needed. Several systems that are able to solve this problem are already on the market (e.g. Ubisense[40], Lukotronic[41] or the "Flock Of Birds"[42] system). The key disadvantages of these systems are that they are very expensive (e.g. the price of the

[40] www.ubisense.net (last accessed on 2013/06/18)
[41] www.lukotronic.com (last accessed on 2013/06/18)
[42] http://www.ascension-tech.com/realtime/rtflockofbirds.php (last accessed on 2013/08/31)

Ubisens system is about 15.000 €) and in general they involve a considerable installation/configuration effort.

- Wearable cameras: The increasing availability of small, wearable cameras and powerful mobile processing units has opened the way for so called "first person" computer vision approaches. For example, cameras have already been integrated into wristbands (see [MYK+10]) or mounted on the user's chest (see [PR12] [NCLZ11]) in order to identify objects on images. The distance between the user and objects can be approximated from the object size on the image. However, many computer vision approaches are based on a large amount of training data. The approach shown in [PR12] for example uses about 1200 images per object category in order to reliably detect activities of daily life. This fact would again prohibit the idea of minimized training data sets (that can be collected by the user through easy to perform one-time measurements), which was one of the main requirements of this work.

However, camera sensors have the following advantages compared to the other systems mentioned with respect to real-life applications: First, affordable sensors such as the GoPro camera (see [PR12]) can be used to detect objects on images. Second, so called one-shot training approaches were introduced, that are based on minimized training data sets. Hence, only a few images of each relevant object and a small set of background images have to be taken in order to train the system. Such simple data collection can be done by the user during a one-time environment walk-through. Third, camera systems do not rely on environmental instrumentations.

Based on these considerations, the key idea behind the work described in this chapter is to use a wearable camera in combination with a "one-shot training" procedure. Due to the fact, that only a single image of each object is considered, the problem of differentiating relevant objects from the background becomes more challenging. Consequently, the idea was to extend the camera system with simple model based approaches describing physical preconditions, that generally apply to activities in order to reduce the corresponding search space. The concepts observed were:

1. Simple hand-object location indicators can be used to
   a) identify and exclude signal segments that with a high probability do not belong to related object interactions.
   b) constrain the number of possible objects for remaining segments.
2. Computer vision techniques can be used to
   a) constrain the search space by limiting image scaling. Thus, relevant objects that are located on images, but are too small, are rejected as they do not belong to an object interaction with high probability. This idea is based on the fact that during the one-shot data collection, objects were taken from a similar distance to the camera-object proximity during real interaction events (see Figure 4.9).
   b) reduce false classifications due to the assumption, that in almost all cases only a single object can be located on images during object interaction events (see Figure 4.9, Images (b) and (e)).

In the following, the ideas shown are introduced in detail.

4.5.1.1 Hand-Object Location Indicator: Proximity

As a first step, indicators for the hand-object location need to be determined. Interactions considered are based on hand-object contacts, whereby that the distance between the user's hand and the objects is quite small or even zero during object interactions. Thus, the most obvious indicator is proximity. As was already mentioned, several concepts can be used to determine a "small" hand-object distance (e.g. RFID systems or high precision trackers). However, a rather

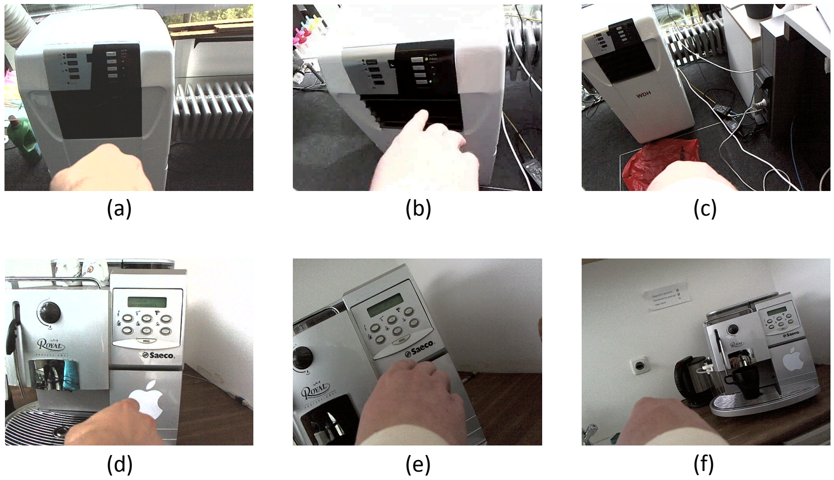


Figure 4.9: Coffee machine and air conditioner: Images taken by a wrist worn camera during a one-shot training data collection (Image (a) and (d)), during real object interaction events (Image (b) and (e)) and while performing background activities (Image (c) and (f)).

simple and low-cost model-based solution, that does not rely on large amounts of training data, is the use of a wrist mounted infrared proximity sensor. Such sensors deliver quite accurate distance information (centimeter range). Simple distance thresholds can be used to define a "small" hand-object distance. Such an approach was followed in this work. Proximity information is used to reject signal segments in which the user's hand is definitely not in close vicinity to an object. The search space can be reduced in this way.

The following algorithms and processing steps were applied:

- Hand-object distance calculation based on a wrist-mounted proximity sensor. The system delivers binary proximity information, which indicates if the current hand-object distance is too large to belong to an object interaction event or not.

4.5.1.2 Hand-Object Location Indicator: Vertical Hand Position

The remaining segments may be related to object interactions. However, the system has no indication of what kind of object the user is interacting with. Hence, the following idea was used to select relevant object candidates and to reduce the number of possible objects: Many objects are located in a fixed place within their environment (e.g. coffee machine, printer or light and shutter switches). Consequently, users have almost the same posture while interacting with a specific device (e.g. bending down in case of objects that are located close to the ground). With this in mind, the current posture of a person during a specific object interaction depends on the object and can be used to limit the number of possible object candidates.

In order to determine human posture, complex three-dimensional human body models can be created. However, the most important part of the body model with respect to object interaction events, is the height of user's hand. Due to the fact, that objects can be described by their vertical position, simple hand-object height comparisons can be used to limit the amount of relevant object candidates. This feature can be easily derived based on inertial sensors and simple body-part measurements only. Although several devices may be placed in almost the same vertical location, this feature is powerful enough to filter out objects that are definitely not located on the same level as the user's hand. Hence, the search space can be reduced and a pre-selection of possible objects can be made. Due to the fact, that the height of objects can be determined by simple one-time measurements, even an occasional repositioning of items is not a problem. However, the approach shown will need unacceptable re-configuration efforts in the case of frequently moved items like food, cutlery or glasses as they have been considered in [LPB+10].

The following algorithms and processing steps were applied:

- One-time measurements of human body-parts and the height of items available.
- Calculation of the user's hand height within a spotted interaction event based on a simple human body model.
- Comparison of the user's hand height with the pre-defined list of item heights.
- Selection of object candidates by taking inaccuracies of both measurements into account.

Besides the vertical hand position, an obvious idea might be to analyze the horizontal hand position as well. This concept and the related problems are discussed later in Section 4.5.3.

4.5.1.3 Computer Vision Based Object Recognition

As a next step, the remaining segments and the pre-selected set of possible objects were used to identify the object the user is interacting with. Therefore, images were captured by a wrist-mounted camera. Images, that correspond to the spotted segment were processed using a vision-based object identification approach. Several methods and features were widely used by the research community to identify objects in images. Examples are color features (e.g. [MYK+10]), SIFT-based features (see [Low99] [PWF09]) or HOG features (see [DT05] [LLZ+11] [ZZS07] [Dal06] [EZW+06]). Due to the fact that color features are known to be very sensitive to changing lighting conditions, which is usually the case in real-life environments, they have not been considered in this work. A SIFT-based object recognition system developed by David Lowe[43] was applied on sample images of objects considered in this work during an initial experiment. It turns out, that this standard implementation was not able to reliably recognize objects of interest. The reason for this is, that many of the considered objects in this work are quite texture-less (e.g. "Light Switch", "Microwave" or "Cupboard") and hence, they have little visual information on their surfaces. This fact makes it difficult to use local feature descriptors like SIFT. The problem of texture-less objects and SIFT was also discussed in [KHP07], [HHC+11] and [HCI+12]. Amongst others, objects such as a microwave, personal computer and refrigerators were considered. Besides, the immense number of necessary key points has limited the usage of standard SIFT implementations in object recognition applications (see [PWF09]). By contrast, a HOG based approach delivered promising results. Consequently, it was used in this work to identify objects within spotted segments.

The following algorithms and processing steps were applied:

- Recognizing objects on images based on HOG features, object models (SVMs trained by using a minimal training set) and a pre-selected list of feasible objects.
- Handling multiple object findings on a single image.
- Determining the winning object and rejecting low-ranked objects.

The multi-modal system proposed is called the *core on-body system*.

[43]http://www.cs.ubc.ca/ lowe/keypoints/ (last accessed on 2013/09/16)

4.5.1.4 Core On-Body System: Training and Configuration Effort

At this point, it is worth highlighting again, that the system proposed is purely based on simple one-time measurements and easy-to-perform data collections. In the following, the necessary training effort and related data collection steps are summarized:

- One-time human body and proximity sensor placement measurements.
- One-time determination of items' vertical position.
- One-time image capturing of available objects (one-shot capturing: A single image per object is sufficient).
- Labeling of objects/significant object areas in images.
- Artificial generation of a training data set by adapting captured images (gamma filter).
- Background data collection (Recording environment images during a one-time environment walk-through is sufficient).
- Creation of object models based on HOG features, SVMs and collected training data.

The defined research questions are to a large part related to fusion approaches based on the core on-body system. Hence, the general key fusion concepts and ideas are addressed in the following, before the next research question is discussed.

4.5.1.5 What kind of Fusion Concepts can be used to improve the Core On-Body System?

The core on-body system has already realized ideas and concepts, that aim to reduce the search space of the considered problem. Based on these ideas and further considerations related to the system's structure, the following three aspects can be determined:

- Firstly, the search space can be significantly constrained, if the amount of spotted intervals, that were chosen based on the hand-object proximity, can be further reduced. Therefore, features may be used, which are able to filter out segments that are not of interest (e.g. interval duration). This concept is hereafter referred to as SSR-I.
- Secondly, the list of possible objects, that were assigned to spotted intervals based on object and hand height comparisons, can be further constrained. Therefore, additional object related pre-conditions can be taken into account such as the location of the user (as it is implausible that the user is interacting with a device while not being in close vicinity to it) or measurable interaction effects on objects. This concept is from now on referred to as SSR-II.
- Thirdly, the amount of processed images related to spotted intervals can be reduced. The key idea is, to integrate systems, that are able to indicate and to sort out images, that definitely do not belong to relevant object interaction events. This concept is hereafter referred to as SSR-III.

4.5.2 Second Research Question: Impact of Location and Device Use-Mode Detection Methods that rely on Minimal Training Data

Section 4.2 showed, that a large amount of activity recognition approaches exist in general, that may be an appropriate extension of the core on-body system shown. However, it was already discussed at the beginning of this section, that the activity types observed are determined by both a set of preconditions and the effects on the objects involved. Considering these aspects, topics already covered by the thesis and aspects about the possible search space reductions SSR-I - SSR-III (see Section 4.5.1.5), the use of the following system concepts is obvious:

- Location systems: The core on-body system so far only uses the information that the user is "close" to an unidentified group of objects, which are located on a similar vertical plane as the hand. In contrast, combining this information with more precise location information could be very useful to further narrowing down the search space. For example, objects that are placed at a similar height to the users' hand, but are located far away from the current user position can be discarded straight away. This idea is related to SSR-II.

- Device monitoring systems: As was already mentioned, the activities considered have effects on objects, which result in many cases in a use-mode change (considering for example electronic devices) or motions (e.g. opening/closing a cabinet door or picking something up). Consequently, objects that show no change (regardless of the type of change) can be excluded from the list of possible objects. This idea is related to SSR-II.

These concepts and related approaches are discussed in greater detail as follows.

4.5.2.1 Recognition of Device Operating Modes

As a first step, the aspect of device monitoring is discussed. In Chapter 3, the thesis already introduced two methods to recognize operating modes of common electronic devices and water taps in an easy, unobtrusive way and based on minimal training effort. Consequently, both systems are used again in this chapter to detect use-mode changes. As already described, objects selected by the core on-body system as possible interaction candidates (of a spotted segment) were rejected, if they haven't changed their use-mode during the spotted event. Note again, that it is not possible to detect use-mode changes for all considered devices on the basis of the systems introduced in Chapter 3. Thus, the remaining objects like cabinets or ring binders could not profit from this approach. However, the use of further object instrumentations (e.g. wireless acceleration sensors on ring binders to detect motion) were not considered in this work as it would destroy the idea of unobtrusive and minimal environmental instrumentation.

The following algorithms, processing and training steps were applied:

- A simple, one-time deployment procedure. Systems described in Chapter 3 can almost be used out-of-the-box. In theory, power profiles of electronic devices can be delivered by service providers. The water approximation system is already easy to deploy and also to set up. As a consequence, it was granted a "Best commercial potential" award.

- Configuration effort: Register monitored objects in the system.

4.5.2.2 Sub-Room Level Activity Monitoring

With respect to location, the thesis already introduced a sub-room level positioning system in Chapter 2. The concept of pre-defined regions of interest (ROI) tagging objects or even groups of objects located next to each other is considered again in this section. Instead of using a standard object tracking module, simple vision-based motion detection algorithms (combined with ceiling cameras) are applied in this chapter in order to recognize unidentified activities within ROIs. Since the focus is on a multi-user scenario, detected ROI activities and therefore the included objects/groups of objects cannot be assigned to the monitored user. Besides, the information that some activity was recognized in the close surroundings of an object, does not imply an object interaction (e.g. people talking to each other while standing in front of the coffee machine). However, the proposed system is easy to deploy and provides information, that is powerful enough to reduce the search space of the core on-body system as objects included in inactive regions of interest (ROIs without recognized activity) can be rejected with a high probability.

The following algorithms, processing and training steps were applied:

- A simple vision based motion detection algorithm applied to pre-defined sub-room regions. Such algorithms are already included in many commercial cameras (e.g. Axis 212 PTZ surveillance camera[44]).
- Configuration effort: Register monitored objects in the system and define regions of interest for objects or groups of objects.

4.5.2.3 Room Level User Location

As already mentioned, high precision tracker systems were not considered in this work due to the previously presented reasons. As shown in Section 4.2, various RF-based approaches exist to detect the rough location of a person. Such an approach was considered in this work in order to locate the monitored user on a room level. Hence, object candidates of a spotted segment were rejected, if they were not located in the same room as the user. The main intention was to evaluate "the price that has to be paid" for focusing on a rough, room level position instead of a sub-room level motion detection system. In contrast to systems already shown, this approach relies on a statistically significant training data set. However, the necessary reference data can be easily collected by the user during a one-time, quick environment walk-through.

The following algorithms, processing and training steps were applied:

- A standard fingerprinting approach was used in combination with a wearable scanner device and RF beacons placed in each monitored room.
- Training effort: Simple one-time collection of reference data (environment walk-through) and one-time object-room assignments.

Besides the consideration of user location systems and effects on devices, further fusion approaches aiming at search space reductions defined by SSR-I – SSR-III are investigated in the following.

4.5.3 Third Research Question: What other Systems describing common Pre-Conditions of the considered Activities and based on Low Training Effort can be used to Improve the Core On-Body System?

In the following, further fusion approaches aiming at improving the spotting performance of the core on-body system by considering search space reduction concepts SSR-I – SSR-III are shown. Systems related to the first key aspect, which was defined at the beginning of Section 4.5, are introduced. Furthermore, the considered systems are based on simple threshold configurations and minimal training data collections.

The main question is: ***What other pre-conditions are valid for the considered activities?*** So far, only aspects of hand-object proximity and the vertical hand-object distance were taken into account. The following ideas were addressed:

- Firstly, the most obvious fact is related to *hand movements*. Each object interaction performed is caused by specific hand activity. Such problems are widely solved by state-of-the-art motion systems using large, statistically significant training sets. However, this would prevent the idea of minimal training data. Besides the recognition of specific activities, hand motions describing interaction events in general can be considered. The idea is, that the user has to raise and/or move his hand forward immediately before an interaction starts. This results in a significant hand acceleration that can be measured by a wrist mounted motion sensor. Similar model-based concepts were introduced in [ZS08] and [ZWS09]. Short but fixed hand positions and turning points of hand movements were used to split a continuous data stream in segments of interest. Based on relevant

[44] User's manual: http://www.axis.com/files/manuals/um_212ptz-v_38187_en_1001.pdf

segments, gestures were classified using information about the direction of movements between turning points as well as the shape of turning points.

- Secondly, almost all considered object interactions are performed while the *user is standing in front of the object*. State-of-the-art modes of locomotion recognition or even simple standing detection systems can be used to detect such events.

- Thirdly, in almost all cases the user's hand has to *point towards the object*. This implies both a specific vertical and horizontal orientation. Whereas the vertical position of the user's hand has already been considered by the core-system, the determination of the horizontal orientation is much more difficult for the following reasons:

  - People can perform object interactions from different positions. This implies that the horizontal heading might be completely different between two interaction events. For example, pushing a light switch while standing in front of the object implies, that the user's arm points straight forward. In contrast, when pushing it while standing nearby and talking to a person, the hand can point in a completely different direction (with respect to the body of the person). Obviously, this assumption is not valid for the vertical position.
  - Inertial sensors are widely used to calculate the heading of objects (in this context the user's hand) based on magnetic field measurements. However, the magnetic field can be easily disturbed by materials included in the environment or by electronic devices. Consequently, the calculated orientation can be afflicted with errors.

  For these reasons, a reliable determination of the user's hand heading based on magnetic field measurements is difficult. However, this work makes use of the problems mentioned. The idea is, that magnetic field disturbances in the close surroundings of objects may be characteristic for a specific object. Hence, object specific magnetic field signatures can be determined and based on this information, objects with totally different signatures than the current measurement can be rejected with high probability. This way, the problem of magnetic field disturbances is used as a feature.

- Fourthly, object interactions have a *limited duration*. This information can be easily derived from spotted segments and can be used to reject segments lasting an inappropriate time.

Systems described in the following were selected based on these considerations.

4.5.3.1 Hand Movement Intensity

This work distinguishes between two intensity levels: low and high hand movement intensities. Based on the considerations made, images that were captured during slow hand movements are rejected with a high probability. The idea is, that such images may belong to periods where the hand does not move towards the object or is very close to the object (during the moment the user touches a button for example). Hence, the camera is not able to capture the object in both cases (in the latter case the object may be occluded by the user's hand or the camera is located too close to the object) and associated images need not be analyzed. This concept is related to SSR-III.

The following algorithms, processing and training steps were applied:

- Determining hand movement intensity.

- Rejecting images captured during low hand movement intensities.

- Configuration effort: The system is simply configured by a single intensity threshold used to differentiate between low and high movement intensities. Such a threshold may be pre-defined in a user-independent way or by the user himself through simple intensity measurements.

4.5.3.2 Mode of Locomotion

Indisputably, the most widely used approach to recognize modes of locomotion is based on on-body motion sensors (e.g. [RM00] [VLC00] [LM02] [MHS01]). As was already discussed, a large amount of considered object interactions are performed while the user is standing in front of the object. Consequently, a simple approach based on on-body motion sensors was used in this work to detect periods, in which the user is standing still. Images of spotted segments, that were not captured during standing periods, were not processed as they do not include objects of interest with a high probability. This concept is related to SSR-III.

The following algorithms, processing and training steps were applied:

- Determining a user's mode of locomotion (standing detection).
- Reject images (included in spotted segments) that were captured during walking periods.
- Configuration effort: Standing detection systems based on motion sensors can be easily configured or pre-trained in labs in a user-independent way.

4.5.3.3 Magnetic Field Characteristics

This approach is based on the measurement of magnetic field characteristics in the close vicinity of each object. Therefore, object-related reference data has to be collected once. The data set consists of magnetic field measurements of a wrist-mounted inertial sensor and was recorded while the user pointed towards specific objects and their close surroundings. This way, object candidates that do not fit to the current magnetic field signature (in combination with time duration features) within a spotted segment, can be rejected. Note, that object-related reference data, which was used in this work, was recorded about four weeks before the experiment took place. The data recording was performed within a real research lab at the University of Passau, Germany. Due to normal working conditions, small environmental changes were performed by students and employees during that time (e.g. moving furniture) and therefore the system has to face small magnetic field changes as is usually the case in real-life scenarios. This concept is related to SSR-II.

The following algorithms, processing and training steps were applied:

- Using a fingerprint-based method to compare object-specific reference data with magnetic measurements of a wrist-mounted inertial sensor.
- Determining the duration of an object-related pointing event within a spotted segment.
- Reject objects that do not fulfill both requirements of heading and duration.
- Training effort: Collection of object-related reference data. This can be done by one-time, object-specific measurements. Besides, an object-independent time threshold must be defined once.

4.5.3.4 Time-Related Features

It is obvious that object interactions have a limited duration. However, the interaction duration depends very much on the activity performed (e.g. pushing a button is less time-consuming than preparing a cup of coffee). Besides, the duration of an action can vary widely from one way of doing it to another and can also be user-dependent. This fact in combination with the large amount of possible activities that can be present in a real world scenario makes a system pre-training in labs difficult. Due to these considerations, the definition of object-related time thresholds was not considered. In contrast, the idea is to disregard spotted segments, that with a high probability do not belong to an object interaction. An example is that users lean their arm on the table while talking. During such activities, the arm-table distance is quite low and a segment is spotted. As the duration of such segments is much longer than for almost every object interaction, they can be filtered out. This concept is related to SSR-I.

The following algorithms, processing and training steps were applied:

- Calculating time durations of spotted segments and rejecting segments with too long durations (using a simple threshold-based approach).
- Configuration effort: One-time configuration of a single time-threshold.

The next research question deals with the impact of a state-of-the-art motion approach.

4.5.4 Fourth Research Question: Impact of a State-Of-The-Art Activity Recognition System using a Large Training Data Set

In the following, restrictions on the use of statistically significant and large training data sets were temporarily softened. The key idea is, that detailed information about performed gestures/hand activities can be assigned to specific object interactions. Due to the fact that subtle and similar hand activities were focused on in this work (e.g. pushing buttons on several devices), such systems may not be able to solve the spotting problem reliably. However, the idea is that they may be able to reduce the search space of the core on-body system (e.g. a push button activity is definitely not related to objects like a cabinet or water tap). They can therefore contribute to a performance improvement of the on-body core system.

In Section 4.3.1 it was already shown, that on-body motion sensors are not able to reliably spot activities considered in this work alone. However, the key idea of this section is to combine results of both systems: the core on-body system and the motion sensor system. Section 4.7.1 will show, that the inertial system itself is able to reach a higher recall value than the core on-body system. However its corresponding precision is significantly smaller. Consequently, the idea was to improve the recall of the core on-body system by choosing the winner object not only on its determined recognition score, but also by using information about the results of the inertial system.

The procedure is as follows: Both systems are able to deliver a ranked list containing possible winners for spotted segments ordered by recognition probabilities. The first winner candidate of the core on-body system (from the ranked list of possible winner objects) that was also included in the list of possible winners of the inertial system was chosen as the final winner. If matching objects could not be found, the original result of the core on-body system has been taken. The proposed fusion procedure is related to SSR-II.

The following algorithms, processing and training steps were applied:

- Spotting segments using the core on-body system.
- Spotting segments using the inertial system.
- Determining ranked lists of winner candidates for both systems.
- Choosing the highest ranked object of the core on-body system that is also included in the list of possible winner candidates of the inertial system as final winner.
- Training effort: The training and configuration procedure of the core on-body system has already been described above. The inertial system has to be trained on a user-independent data set containing several repetitions for each activity. In this work, 15 repetitions were performed per activity.

Section 4.7.7 will show that the motion system is not able to provide enough information to improve the results achieved by the fused core on-body system.

4.5.5 Fifth Research Question: Impact of Replacing the Wearable Camera with a State-Of-The-Art Motion System

This research question is motivated by the following two facts:

- Vision-based object recognition methods are known to be computationally expensive. A SIFT based object recognition approach introduced in [Low99] needs almost two seconds to process a single image, which is too slow with respect to wrist worn cameras and fast hand movements. A HOG based approach described in [DT05] needs less than one second for images with 320x240 pixels. Due to the fact, that this work considers camera images with a resolution of 640x480 pixels, the performance of a standard HOG approach may not be sufficient for real-time evaluations. However, optimized HOG systems were introduced, that are able to recognize objects on images in real time. One example is shown in [ZZS07]. About 25 frames per second can be processed with a resolution of 320x240 pixels (a Xeon Dual-processor 3.6GHZ CPU was used).
- Compared to vision-based approaches, motion systems are much less computationally expensive. Motion data can even be processed on low power processors (as are included in smartphones) with about 100Hz (e.g. Xsens inertial system).

We have seen, that the state-of-the-art motion system definitely is neither able to provide sufficient information to reliably spot considered activities nor to improve the results achieved by the fused core on-body system. However, the question is, if the core on-body system can achieve the same or at least similar results when simply replacing the wearable camera with the motion system. Therefore, the following two concepts were considered:

4.5.5.1 Replacing the Wearable Camera and the Proximity Sensor with a Motion System

First, a reduced sensor setup is considered. Besides replacing the wearable camera, the proximity sensor and hence information about the hand-object distance was not considered. The resulting system consists of inertial sensors only, but uses hand height information derived from a human body model in addition to the system described in Section 4.3.1.

The idea is, that the user's hand height is quite constant (even if it is only for a short period) during an object interaction. Consequently, segments in which the hand height variation is less than 10 cm were spotted as interesting intervals. Based on the average hand height and a pre-defined list of object heights, relevant object candidates were selected. As a next step, an inertial sensor system was applied and the object with the highest score was chosen as final winner. Finally, a threshold was applied to reject low ranked winners.

Algorithms and Processing Steps:

- Spotting of interesting intervals using information about the user's hand height variation.
- Determining a list of feasible object candidates based on the average hand height and a pre-defined list of object heights (the same approach has been chosen for the core system).
- Applying trained inertial sensor systems for each object candidate on spotted segments.
- Determining the final winner and rejecting low ranked winners.

4.5.5.2 Replacing the Wearable Camera with a Motion System

The idea is to use exactly the same methods as have been used by the core on-body system, but replacing the camera sensor with a motion system. Consequently, the inertial sensor system was used to recognize object interactions on top of the spotted intervals and based on a pre-selected list of possible object candidates instead of the introduced computer-vision-based recognition system.

Algorithms and Processing Steps:

- Spotting of interesting intervals using hand-object proximity information (the same approach was chosen for the core system).
- Determining a list of feasible object candidates based on the average hand height and a pre-defined list of object heights (the same approach was chosen for the core system).
- Applying a trained inertial sensor system on object candidates of spotted segments.
- Determining the final winner and rejecting low ranked winners.

Section 4.7.9 will show that the proposed fusion techniques achieve significantly worse results than the core on-body system. Consequently, further considerations and evaluations were not performed on this topic.

4.5.6 Sixth Research Question: Impact of Fusing Multiple Systems that do rely on Minimal Training Data

So far, only fusion approaches with single systems have been considered. The next obvious step is to evaluate the impact of fusing multiple systems with the core on-body system. As Section 4.7.3 will show, the fusion approach based on the intensity of the user's hand movement is unable to improve the spotting performance of the core on-body system. In contrast, even slightly worse results have been achieved. Consequently, this approach will not be considered any further.

4.5.6.1 Impact of Fusing Location Systems

The idea is, that the combination of a room-level location system and a sub-room level activity monitoring system may result in a significant reduction of search space. However, the spotting problem is not completely solved as several people can still be located in the same sub-room region as the monitored user (e.g. people standing close together around the coffee machine). Consequently, the system is not able to determine who is responsible for an object interaction. Besides, the information about magnetic signatures of objects is related to their location. Hence, it was also evaluated in combination with the other location techniques. Specifically, the core on-body system has been fused with:

- A room level location and a sub-room level activity recognition system.
- A room level location system and the magnetic signature of objects.
- A sub-room level activity recognition system and the magnetic signature of objects.
- A room level location, a sub-room level activity recognition system and the magnetic signature of objects.

4.5.6.2 Impact of further Fusing Approaches based on the Core On-Body System and Location Approaches

Based on the results achieved by location systems (Section 4.7.6 will show, that the combination of a sub-room level activity recognition system and the magnetic signature of objects performs best) several fusion approaches were investigated. The main idea was to evaluate the "price that has to be paid" for minimal instrumentation and sensor equipment. The following fusion approaches with the core on-body system were focused on:

- Fusing all selected systems: Time features, modes of locomotion detection, device operating mode detection, a sub-room level activity recognition system and the magnetic signature of objects.
- A reduced fusion approach based on location and device operating mode information: Device operating mode detection, a sub-room level activity recognition system and the magnetic signature of objects.
- A fusion approach that does not rely on device use-mode information[45]: Time features, modes of locomotion detection, a sub-room level activity recognition system and the magnetic signature of objects.
- A fusion approach which is completely independent of environmental instrumentations: Time features, modes of locomotion detection and the magnetic signature of objects.

Section 4.7.6 will show, that the fusion of all selected systems provides almost the same spotting performance as the reduced fusion approach. Hence, it was not investigated in the following.

The performed evaluations were based on recall, precision and EER. However, aspects of fragmentations, merges or under- and overlays were not considered. Jamie Ward considers this problem in [WLG11] and introduced two evaluation procedures based on frames and events. Both evaluation procedures were applied on the remaining three fusion approaches in order to achieve more detailed evaluation results.

As a next step, the recognition of concrete subtle hand activities which build the basis of the spotted object interaction events was considered.

4.5.7 Seventh Research Question: The Recognition of Subtle, Barely Distinguishable Hand Activities

The final research question observed in this chapter deals with the topic of spotting and recognizing subtle, barely distinguishable hand activities describing the performed object interactions in greater detail. Within the third research question and in Section 4.2 several state-of-the-art methods recognizing arm gestures in general were introduced. However, we saw that the fused core on-body system was able to reach a reasonable performance when spotting object interactions. As it was one of the key objectives to minimize environmental instrumentations as well as configuration and training effort, fusion approaches based on additional sensor modalities were not investigated further. In contrast, the main consideration was:

How can the fused core on-body system be used to derive subtle, hardly distinguishable hand activities based on identified object interactions?

Section 4.7.8 will show, that the best spotting performance was achieved by the core on-body system fused with an device operating mode detection system, a sub-room level activity recognition system and the magnetic signature of objects. Consequently, this configuration was chosen to recognize subtle hand activities. Based on the main components included, the following considerations were made:

- Camera systems: One possible solution to detect subtle hand activities would be to spot and track the user's hand (e.g. [RK94] [ZUA02]). Such methods can be used to determine what part of a device the user has touched and hence to derive the performed interaction type. Another approach (see [GWM11]) shows how inertial sensors and wearable cameras detecting the user's gaze and analyzing the surrounding scene can be fused to identify concrete object interactions and guide the user in industrial applications. Camera-based

[45]This decision was based on the consideration that common objects are unable to provide the necessary information to a large extent – they would therefore have to be equipped with additional sensors.

workflow detection systems were also introduced to detect activities related to objects. In [PS12] amongst others, a printing device and corresponding activities such as "touch", "open upper lid" and "remove cartridge" were considered. Obviously, detailed information about specific devices or arm gestures must be available to train and realize such systems. Due to this fact, further image processing approaches were not considered in this work.

- Inertial systems: We have already seen, that a motion system does not provide enough information to spot object interactions reliably. However, they may include enough information to distinguish between a finite set of possible object-related activities. For example, distinguishing between actions like "open microwave", "close microwave", "turn microwave on" or "clean microwave" is much simpler than spotting these activities within a large amount of background data. However, it may also be a difficult task for the system to differentiate barely distinguishable activities like turning on different devices.

- Device operating modes: Operating modes of devices emerge from specific user activities. For example, the fact that a device is turned on results clearly from a user activity like "push device-on button". Hence, operating modes can be mapped to specific arm motions. As was already mentioned, not all objects can be easily instrumented in an unobtrusive way. Consequently, only a reduced set of arm gestures can be detected by this approach.

Based on these facts, three concepts were investigated (see Figure 4.10).

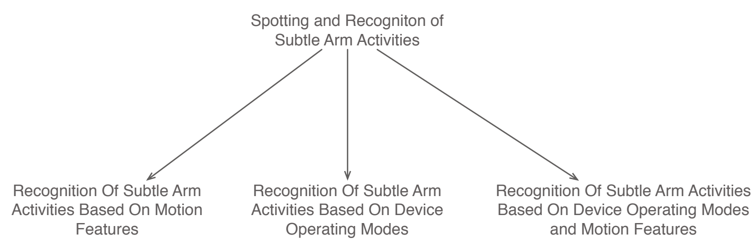


Figure 4.10: Overview: Spotting and identification of subtle arm actions.

4.5.7.1 Subtle Arm Activity Recognition based on Motion Features

Motion systems were used to distinguish between a finite set of object related hand activities only. The key idea was to identify the kind of object with which the user interacts in a first step using the fused core on-body system. Motion systems pre-trained for each object and able to differentiate between object related activities were applied to each identified interaction event. To make the training procedure simpler, background activities were not considered. Consequently, the resulting system is not able to reject falsely classified object interaction events. In contrast, it provides more detailed information about the object interaction by recognizing the underlying hand activity. A big disadvantage of this approach is, that it relies on a statistically significant and large training data set for each activity observed.

Algorithms and Processing Steps:

- Spotting and identifying object interactions using a fused core on-body system.
- Applying an object-specific inertial system on motion data of the object interaction event identified.
- Determining the subtle hand activity performed.

Some activities cannot be distinguished by motion analysis (e.g. turning a device on or off). Consequently, such motions have been joined together.

4.5.7.2 Subtle Arm Activity Recognition based on Device Operating Modes

Another way to recognize the kind of object interaction performed is to map information from operating mode changes to arm activities. We have already seen in Chapter 3 that operating modes and device states of electronic devices and water taps can be identified reliably. Based on an identified object interaction event, the concrete operating mode change, that was performed within the spotted interaction segment, was analyzed and mapped to a specific arm activity. A big disadvantage of this procedure is, that all devices have to be turned into smart devices (able to provide use-mode information) which is in many scenarios neither practical nor possible. Nevertheless, the impact of such devices should be investigated here. Besides, this solution is based on minimal training data.

Algorithms and Processing Steps:

- Spotting and identifying object interactions using a fused core system.
- Determining the objects' operating mode change.
- Mapping the new operating mode to an arm activity.

Some objects can't be easily turned into smart devices. For such devices, related arm activities have been joined together.

4.5.7.3 Subtle Arm Activity Recognition based on Information from Device Operating Modes and Motion Features

Finally, concepts of mapping operating modes to arm activities and analyzing motion patterns were combined. This way the set of recognizable activities was enlarged. Depending on the considered object either smart appliances (if available) or an inertial system was applied in order to identify the hand activities performed.

Algorithms and Processing Steps:

- Spotting and identifying object interactions using a fused core system.
- Determining the objects' operating mode change and mapping the new operating mode to an arm activity **or**
- Applying an object specific inertial system to motion data of the object interaction event identified and determining the hand activity performed.

As a next step, the applied evaluation procedure is described, followed by a detailed description of a concrete system implementation and in-depth system evaluation.

4.6 Evaluation Procedure

The on-body core system and fusion approaches considered were evaluated following exactly the same procedure as described in Section 4.3.3.

The core system is based on image processing algorithms which can be highly time and power consuming. Consequently, the impact of sensor fusion approaches on the amount of required classification steps and the number of analyzed images were also analyzed. These features were used as performance criteria alongside achieved improvements on recall, precision and EER.

Evaluations were performed "offline" and on the basis of the office data set introduced (see Section 4.3.2). I want to emphasize, that although the data processing and evaluation was performed "offline", all systems except the image-based object recognition were designed to be executable in real time.

4.7 Spotting and Identification of Object Interactions: System Implementation and Evaluation

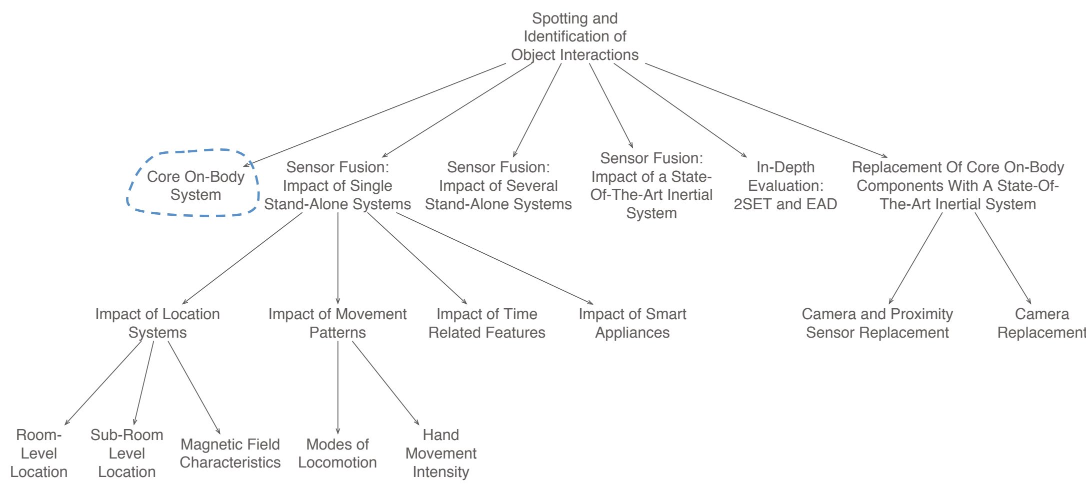
The first part of this section focuses on spotting and recognizing object interactions. Table 4.6 shows the object interactions and subtle hand activities of the office data set again. Altogether, 713 object interactions based on 16 different object types were performed.

Table 4.6: Object interactions and activities of interest grouped by rooms: kitchen, printer room, meeting room and office

Object	Activities (Repetitions)
Microwave	Open (27), Close (27), Start (11), Clean (16)
Coffee Machine	Make Espresso (14), Make Coffee (13)
Power Socket	Connect Cable (27)
Cupboard	Open (27), Close (27)
Wall Cupboard	Open (27), Close (26)
Ethernet Connector	Connect Cable (30)
Water Tap	Fill Big Cup (14), Fill Small Cup (13)
Battery Charger	Put In Empty Battery (15)
	Remove Battery While Charging (14)
Laser Printer	Take Printout (12), Push Button (15)
Ink Printer	Remove Printout (15), Push Button (12)
Climatic Control	Turn Left (13), Turn Right (14)
PC	Turn On (55)
Scanner	On (29), Off (27), Scan Document (27)
Air Conditioner	On (27), Off (27)
Light-Shutter Switch	Light Button (28),
	Shutter Turn Left (8), Shutter Turn Right (19)
Ring Binder	Take Binder (28), Put Binder Back (28)

In the following, the implementation and evaluation of the core system is shown, followed by the realization and evaluation of the sensor fusion approaches focused on.

4.7.1 Core On-Body System



So far we have seen that a mainstream approach based on inertial sensors and trained with a statistically relevant, large training data set is not able to provide a reasonable precision. This section introduces a novel approach based on wearable sensors, which is able to significantly improve the previously accomplished precision value while almost keeping the same recall rate.

4.7.1.1 Sensor Setup

The multi-modal sensor system is based on a basic component (called *BS* in the following) which includes the following sensors:

- A webcam (Logitech C910) mounted on a forearm. A well accepted computer vision algorithm using HOG features (as used in [DT05] [ZZS07]) was utilized to recognize objects on camera images. In that way, objects, the user is currently interacting with, should be identified. Mainstream and low-cost webcams are designed to provide the best image quality while being mounted at a fixed, non-moving location. Consequently, camera parameters were adapted to suit the requirements of a wrist-mounted webcam and the necessity of sharp images during fast hand movements. The following camera parameters were set (using the "libwebcam" library) based on several real-time experiments including fast hand movements and changing light conditions: Brightness = 140; Contrast = 37; Saturation = 26; Autofocus = disabled; Absolute Focus = 68; Absolute Zoom = 1; Sharpness = 72; Auto exposure mode = on; Auto exposure time = 3; Gain = 180. The webcam was configured to provide images with a resolution of 640x480 pixels, and with a frequency of about 15 frames (on average) per second. Price: $<$ 100 €.

- An infrared-based distance sensor (proximity sensor) mounted on top of the webcam. The measured distance between the person's hand and objects was used to spot interesting data segments. The system uses a Sharp GP2D12j0000F proximity sensor formerly manufactured by Toradex[46]. It provides a sampling rate of about 10 Hz and ranges from 10

[46] http://www.toradex.com (last accessed on 2013/06/17)

cm to 80 cm (resolution of 0.001 m). The infrared beam opening is 2 degrees. Price: 30 €.

- Sensors providing orientation information. Sensors are attached to the forearm, upper arm and onto back. On the basis of Euler angles and a body model derived from the simple measurements of rigid human body parts, the height of the user's hand was calculated. The on-body system uses the Xsens inertial system. Price: about 1000 € per sensor (Due to the considerable ease with respect to integration, data acquisition and system setup, the Xsens system was used in this thesis instead of low-cost (below 100 €) alternatives).

Figure 4.11 shows sensor components of BS.

Figure 4.11: Left: Sensor placement on the body (webcam with a proximity sensor mounted on-top and Xsens sensors on forearm, upper arm and back). Right: Images recorded by the wrist-worn webcam.

4.7.1.2 Configuration and Training Procedure

This section explains the training procedure of the basic system in detail. It is shown, that there is no need to collect a large, statistically significant training data set. The system has to be configured and trained only with a minimal training data set.

System Configuration In this context, system configuration means carrying out simple one-time measurements of the environment and individuals. The following measures must be taken to set up the basic system introduced:

- Human body measurement: Length of legs, torso, upper arm and forearm
- Proximity sensor placement: Distance between proximity sensor and middle finger tip
- The vertical position of each relevant object

System Training With a Minimal Data Set Unfortunately, computer vision systems designed for object recognition tasks and based on features such as HOG (see [DT05] [ZZS07] [Dal06]) require hundreds or even better several thousand training images for each object. In the best case the object has to be captured from all possible angles, which is of course not practical in large-scale applications including a large amount of different objects. Consequently, the approach shown only takes ONE training image for each object. A significant area of the object (e.g. control panel) has to be marked once for each object (see Figure 4.12). All in all 80 training images for each object are artificially generated from one captured image by exposure corrections using a gamma filter (see Figure 4.13). Thus, the training amounts of a single action (taking one photo). Additionally, it is assumed that the monitored area is walked through once and random

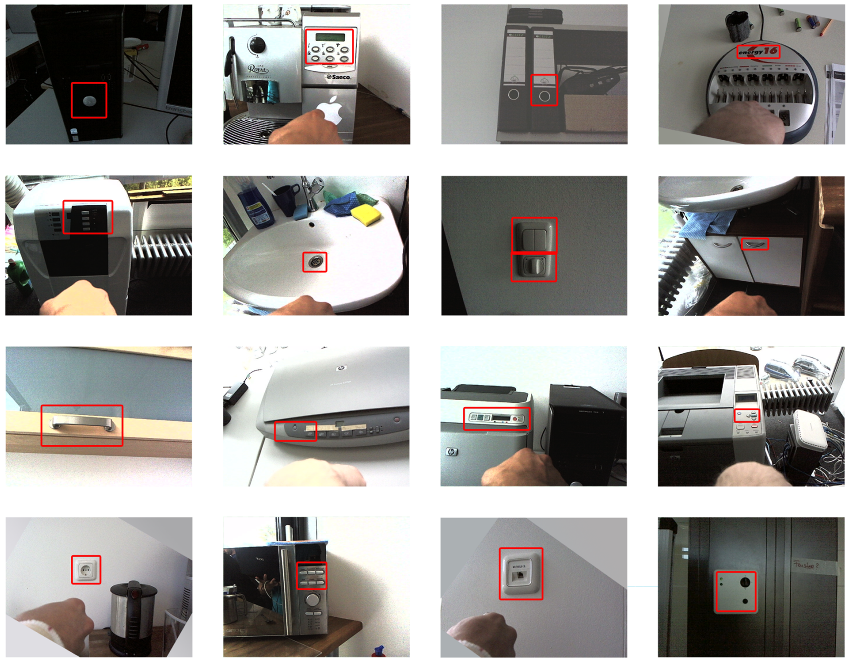


Figure 4.12: Training Images: The red rectangles mark characteristic object areas or the object itself (in case of small objects). Objects from top left to bottom right: PC, coffee machine, ring binder, battery charger, air conditioner, washbasin, light switch, cupboard, wall cupboard, scanner, ink jet printer, laser printer, power socket, microwave, Ethernet connector and climatic control panel.

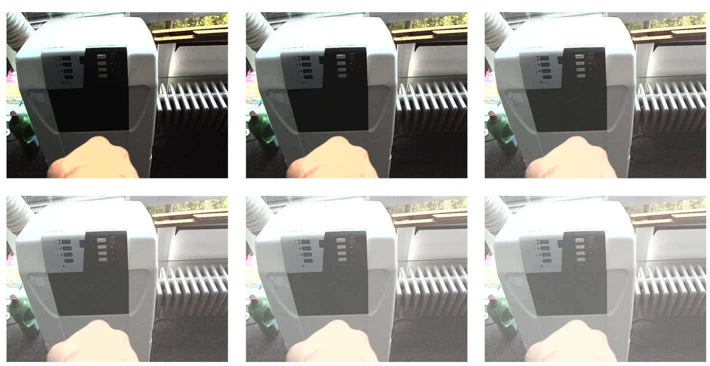


Figure 4.13: Air conditioner: Examples for exposure corrections using a gamma filter.

images are recorded as NULL class examples. In order to solve the object recognition problem concerned, SVMs were combined with HOG features as they have been widely used for similar problems (see [DT05] [ZZS07] [Dal06]). I want to emphasize, that this work does not focus on finding the best suitable computer vision algorithm able to solve the problem in question. On the contrary, the intention of this work was to prove, that the recognition quality of a system which is partly based on a mainstream and well-accepted computer vision approach but using minimal training data (one-shot training) could be significantly improved when combining it with other sensor modalities that are able to restrict the search space considered. Nevertheless, computer vision parameters chosen are briefly introduced in the following: First, a sliding window is used to scan the image with a stride length of 8 pixels. There, the window size is equal to the size of the marked region for each object. Each window is separated in overlapping blocks of 2x2 cells (used for contrast normalization), where each cell has a size of 8x8 pixels. Based on resulting cells HOG features (9 histogram bins and a histogram range of 180) were calculated. Based on artificially created training data and images related to the NULL class (containing background images and training images of other objects), a single SVM (linear kernel) is trained for each object.

4.7.1.3 Segmentation and Classification

Once the system was configured and trained (based on simple one-time measurements and a minimal training data set), the following steps were performed in order to spot and to recognize object interactions:

Spotting interesting time sequences using an infrared distance sensor In general, hand actions involve object manipulation, which means of course that object interactions between humans and objects can only occur when the person's hand is close to an object. Using an infrared proximity sensor, time sequences TS_i were spotted, in which the person's hand is "close" to an object. As it can't be guaranteed that the sensor is always located exactly at the same position and because of the location and structure of objects, a distance close to zero can't always be reached while performing activities. Because of this fact an interesting time sequence is spotted if the distance between an object and the middle finger tip of a person is less than 10 cm[47]. Figure 4.14 visualizes the distance deviation while performing the activity "Connect charger to power socket". This step involves no statistical training and needs only a single measurement related to the proximity sensor placement on the arm.

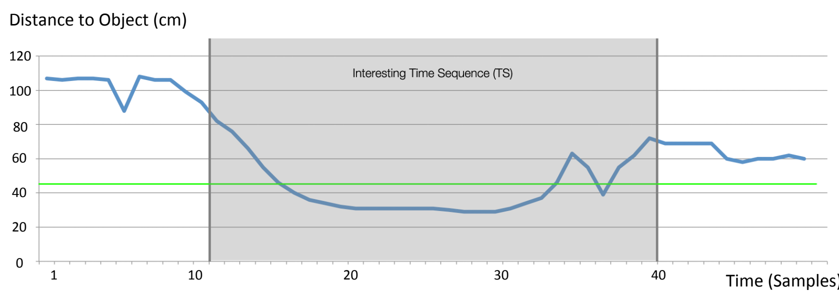


Figure 4.14: Spotting of interesting time sequences: The distance between an object and the user's finger tip must fall below 10 cm to spot interesting time sequences TS_i. The green line marks a distance threshold of 45 cm. In this example, the proximity sensor was mounted 35 cm away from the middle finger tip on the forearm and consequently, the green line visualizes a distance threshold of 10 cm between the finger tip and an object. The grey marked area shows the resulting TS_i defined on a second basis.

Assigning relevant objects to a time sequence A simple body model was defined based on one-time measurements of the human body (in detail: length of leg, torso, upper arm and forearm). In combination with orientation sensors mounted on the upper arm, forearm and the back, the height of the user's right forearm was determined. This information was used to split each TS_i into several sub sequences TSS_{i_j}, where the maximum hand deviation is less than 10 cm. Figure 4.15 shows an example of splitting up a TS_i into several TSS_{i_j}. Consequently, the hand height variation within each sub-sequence is quite constant.

Afterwards a set of relevant objects o was assigned to each TSS_{i_j}. There, a TSS_{i_j} contains an object o if the maximum deviation between the average hand height of TSS_{i_j} and the pre-configured height of the object o is between +/- 30 cm (in the following this value is referenced

[47] This value was chosen based on initial experiments.

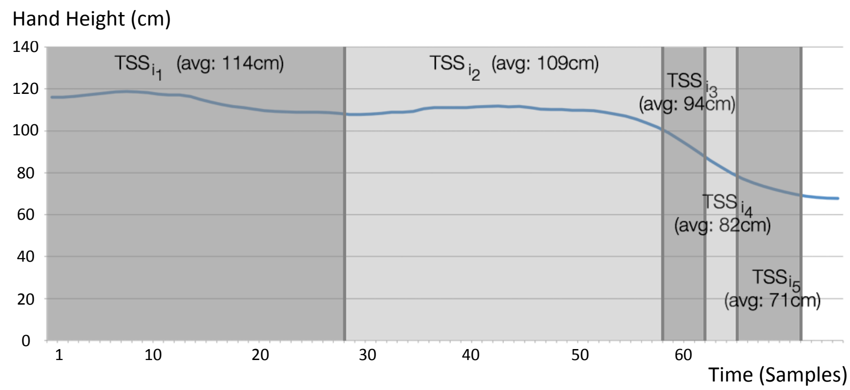


Figure 4.15: The splitting of a TS_i into several TSS_{i_j}. This example is based on the TS_i which was shown in Figure 4.14.

as ($thrDistHO_{up}$, $thrDistHO_{down}$)). This way the inaccuracy of the vertical hand position calculation as well as small hand height variations during object interactions are covered. Figure 4.16 shows the height of each relevant object in the considered scenario. This measurement was performed in a single step without involving statistical training.

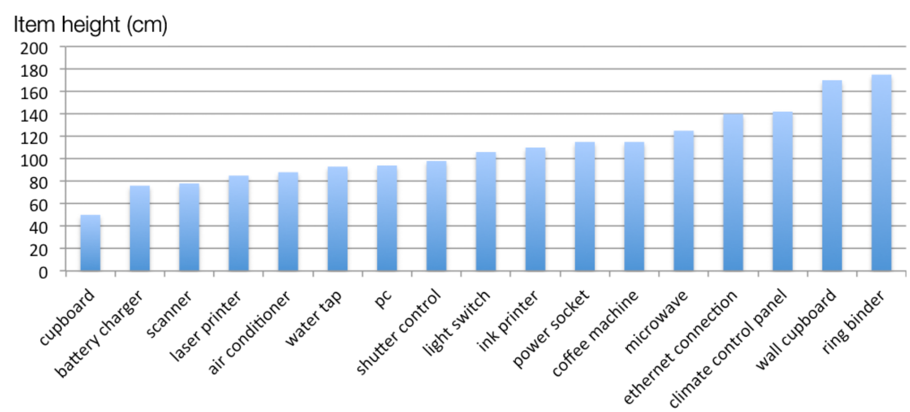


Figure 4.16: Pre-defined object heights. More than half of the objects are located between 80 cm and 120 cm above the ground.

Image based object recognition So far we got several TSS, each containing a list of possible objects. As a next step a standard computer vision approach was used to identify objects on images based on SVM and HOG features. The objective was to recognize objects the user is currently interacting with. The following steps were performed for all images within a TSS:

- Convert the image into greyscale.
- Convert the image using a gamma filter[48].
- Check the images for relevant objects using HOG features and SVMs. Each image is rotated from -90 to +90 degrees (in steps of 10 degrees) and scaled between -0.5 and +4.0

[48]If the average brightness value of the image is below a pre-defined threshold (a value of 70 was chosen as it delivered good results during initial experiments), the image was converted with the help of its average brightness value and a gamma filter.

(in steps of 0.05). This step is needed as the training data set consists of a single training image showing objects in one size and from one angle only. Apart from that, the search space was even more reduced.

- Overlying findings of the same object on one picture are merged using a Non-Maximum-Suppression approach.

Finally, SVMs deliver scored findings on images for each object between -0.4 and 1. It is assumed, that an image can contain only one interesting object, which makes sense when considering the defined requirements on the "hand-object" distance. Consequently, the highest scored object was chosen as winner. To reduce false classifications a threshold on the SVM score (hereafter called $thr_{svmScore}$), that must be exceeded to keep the recognized object, was defined. As the camera is not able to capture the whole object while the person is very close to it, each TSS was extended by +/- 1 second in order to include images captured while the person's hand gets close to the object. Figure 4.17 visualizes the classification procedure again.

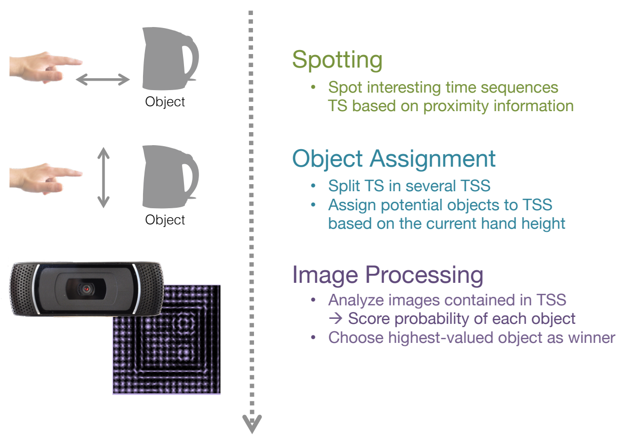


Figure 4.17: BS: Classification Procedure

4.7.1.4 Evaluation

The system evaluation is based on the calculation of $recall$, $precision$ and EER as was described in Section 4.3.3.1. To create $recall$ and $precision$ curves various $thr_{svmScore}$ values ranging from -0.4 to 1.0 in steps of 0.01 were evaluated. For further processing steps the system configuration with the highest $recall$ was chosen due to the fact, that the $precision$ can be significantly improved with the help of additional sensor fusion approaches.

The brown curve in Figure 4.18 visualizes the $(1-precision) - recall$ curves based on different $thr_{svmScore}$ values. The basic system reached an EER of 43%. When looking at the highest recognition rate, a $recall$ of 67% with a corresponding $precision$ of 26% could be achieved using a $thr_{svmScore}$ value of -0.18[49].

As a next step the influence of the pre-defined tolerance value between the current hand height and heights of the objects ($thrDistHO_{up}$, $thrDistHO_{down}$) was evaluated. As a starting point, a deviation of (30 cm, 30 cm)[50] was allowed. The value of $thr_{svmScore}$ was fixed to -0.18 and threshold pair combinations on ($thrDistHO_{up}$, $thrDistHO_{down}$) ranging from (40 cm, 40 cm) to (0 cm, 0 cm) in steps of 5 cm were evaluated. Figure 4.19 visualizes the impact on the $recall$. It can be seen that the highest $recall$ (75%) is achieved by (25 cm, 10 cm).

[49] In the following this value is used to calculate $recall_{obj}$ and $precision_{obj}$ values for systems based on BS.

[50] The value is based on the highest deviation measured during an initial sensor setup phase. Therefore each participant had to point to several markers placed at different heights.

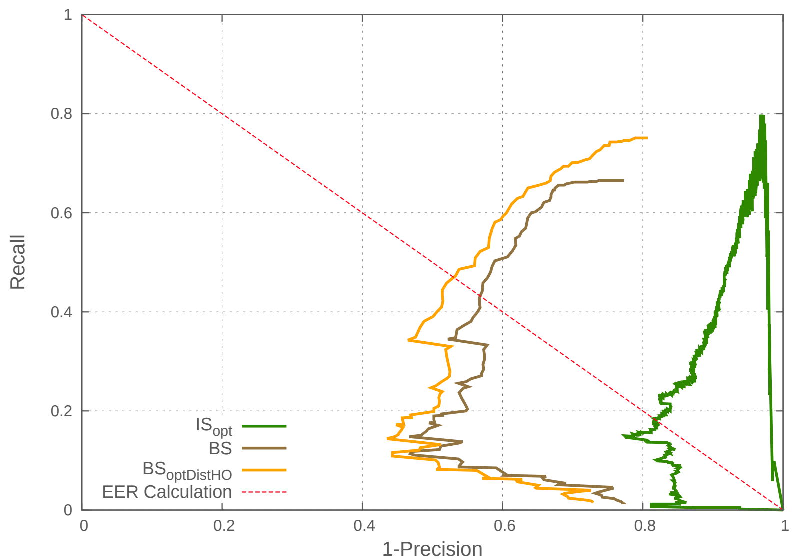


Figure 4.18: $(1-precision)-recall$ curves (based on $thr_{svmScore}$): IS_{opt} (green curve; inertial system), BS (brown curve; basic system) and $BS_{optDistHO}$ (orange curve; optimized basic system).

Unfortunately, the corresponding *precision* loses 4% and falls to 22%. Nevertheless the EER (fixed ($thrDistHO_{up}$, $thrDistHO_{down}$) and optimized $thr_{svmScore}$) could be raised by 4% to 47% (see orange curve on Figure 4.18). In the following, this system configuration is referenced as $BS_{optDistHO}$.

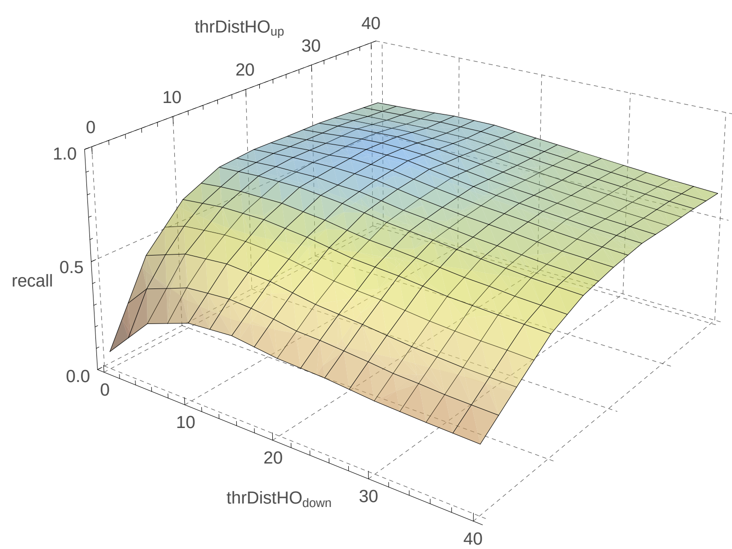


Figure 4.19: Optimization of ($thrDistHO_{up}$, $thrDistHO_{down}$): Impact on *recall*. The highest *recall* (75%) is achieved by (25 cm, 10 cm).

Table 4.7 shows $recall_{obj}$, $precision_{obj}$ as well as EER_{obj} for each object. It can be seen that $recall_{obj}$ values between 43% and 99% can be reached for all objects except the PC. There, none of the related labels were recognized. It seems that the marked region on the corresponding training image (see Figure 4.12) does not contain features powerful enough to identify the PC. In contrast, the $precision_{obj}$ of all objects except "Battery Charger" and "Air conditioner" was very poor. This once again illustrates the difficulty of using a minimal amount of training data in such applications.

Table 4.7: Object overview for $BS_{optDistHO}$: $recall_{obj}$, $precision_{obj}$, EER_{obj} and improvements compared to IS_{opt} in terms of $recall_{obj}$ ($\triangle Rec$), $precision_{obj}$ ($\triangle Prec$) and EER_{obj} ($\triangle EER$).

Object	$recall_{obj}$	$precision_{obj}$	EER_{obj}	$\triangle Rec$	$\triangle Prec$	$\triangle EER$
Battery Charger	83	54	76	-3	52	76
Coffee Machine	85	31	56	11	29	56
PC	0	0	0	-87	-3	0
Air Conditioner	43	60	0	11	56	0
Climatic Control Panel	82	26	41	-18	24	-33
Microwave	99	8	48	39	4	10
Ethernet Connector	73	10	53	-24	8	53
Ring Binder	89	49	70	-6	41	49
Power Socket	70	28	44	-8	27	33
Laser Printer	63	2	19	-22	1	15
Ink Printer	89	5	47	-4	3	36
Light-Shutter Switch	68	7	14	-5	4	0
Scanner	95	7	25	27	4	24
Wall Cupboard	81	14	26	-19	8	-44
Cupboard	78	43	76	-2	39	50
Water Tap	96	5	30	26	3	30
ØAverage	75	22	39	-5	19	22

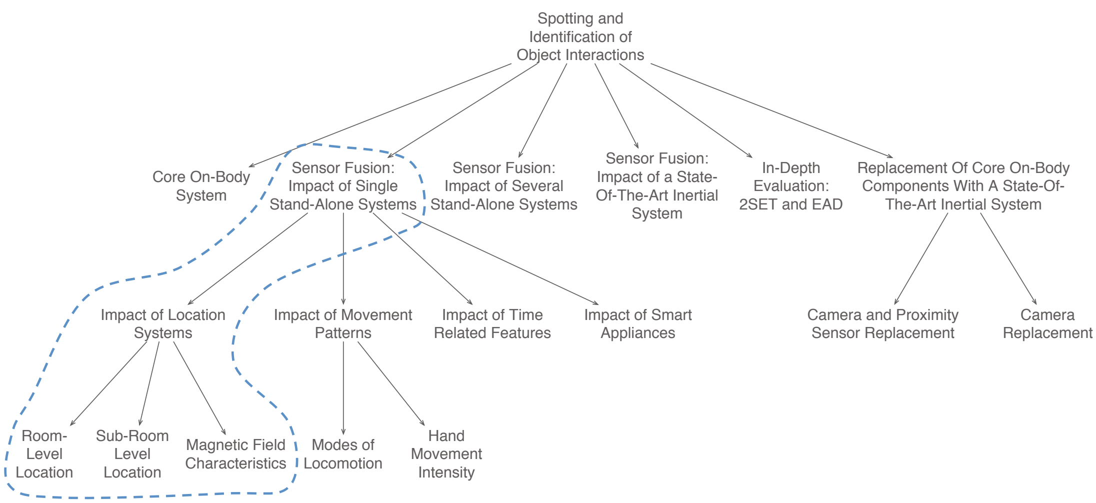
Compared to IS it can be seen, that the average $recall_{obj}$ decreases by 5%. However, the corresponding average $precision_{obj}$ could be significantly increased by 19% and the EER_{obj} by 22%. Having a more detailed look at EER_{obj} and specific objects, it can be seen, that for three objects ("PC", "Air Conditioner" and "Light-Shutter Switch") no improvements could be achieved. For two objects ("Climatic Control Panel" and "Wall Cupboard") $BS_{optDistOH}$ is even worse than IS. Nevertheless, a significant improvement could be achieved for the remaining objects. During the evaluation procedure all in all 205.132 images were analyzed and 1.566.872 classification steps were performed. Table 4.8 summarizes the results for IS and $BS_{optDistOH}$ again.

So far a wearable system based on affordable sensors and a minimal training data set has been introduced, that could deliver a much better recognition quality than a state-of-the-art approach based on inertial sensors and a statistically relevant, large training data set. However, the achieved EER of 47% is still not good enough for many real-world applications. Hence, the next sections introduce several stand-alone sensor systems that have been unobtrusively integrated into the environment or are again worn on the body. In the following, it is shown how $recall$, $precision$ and EER can be significantly improved when reducing the search space by combining the introduced wearable system $BS_{optDistOH}$ with mainstream sensor systems.

Table 4.8: System comparison – $BS_{optDistHO}$: $recall$, $precision$, EER, amount of classification steps (CS_R) as well as analyzed images (AI_R).

System	$recall$	$precision$	EER	CS_R	AI_R
IS	80	3	18	–	–
$BS_{optDistOH}$	75	22	47	1.566.872	205.132

4.7.2 Sensor Fusion – Impact of Single Stand-Alone Systems: Location Systems



The location of a person is rather useful information for activity recognition applications under the condition that activities are always carried out within the same regions. This fits perfectly with the considered scenario as many objects such as "Coffee Machine", "Microwave" or "Water Tap" are normally placed at fixed locations. Consequently, false classifications can be reduced as it is unlikely that an activity such as "Making Coffee" is performed in any room other than the kitchen.

In the following, three location systems, which are based on cameras, RSSI fingerprinting and magnetic field signatures, are introduced. The systems are able to provide different location accuracies ranging from room level to sub-room level as well as a rough location/heading of the user's forearm.

4.7.2.1 A Room-Level Location System

System Setup: In this work, a common approach based on Bluetooth RSSI fingerprinting techniques to locate people on a room level was used. Based on information about the current location of a user the amount of possible objects can be reduced. Therefore, a single Bluetooth beacon was installed in each of the four monitored rooms at random locations. Participants were equipped with mainstream smartphones (see Figure 4.20). A smartphone application was used to scan continuously for reachable Bluetooth beacons and to log them for offline analysis. In the following the combination of the basic system and the Bluetooth based room-level location system is referred to as BS_{BT}.

System Training and Configuration: The training and configuration effort for setting up the proposed system is quite low. Only the following three steps have to be carried out:

- A single Bluetooth beacon has to be mounted at a fixed place in each monitored room.
- Each monitored room has to be walked-through for a few minutes to collect training samples.

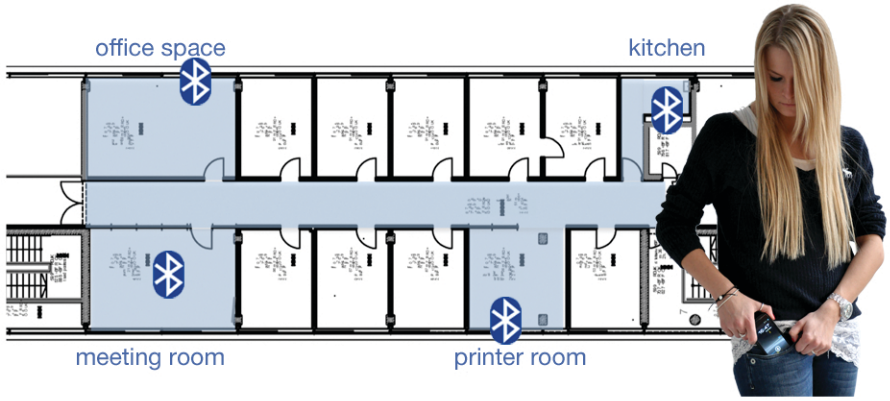


Figure 4.20: Floor map: Placed Bluetooth beacons and participant with smartphone (Samsung Galaxy S).

- Each object has to be assigned to a single or even to several rooms as some objects such as light switches are located in almost all rooms. Table 4.1 (in Section 4.3.2.2) shows relevant objects and their locations.

Classification: The system is based on reference data collected by one-time walk-throughs for each room. The signal strength of reachable and pre-defined Bluetooth beacons is stored in a feature vector v, $v = (sg_{MeetingRoom}, sg_{PrinterRoom}, sg_{Kitchen}, sg_{Office})$. There, the signal strength of beacons that are not reachable is set to -100 dBm. Finally, a set of feature vectors is assigned to each room. In order to localize a user, the system scanned as quickly as possible (the frequency depends on the Android OS; between 0.1 and 1 Hz) for reachable Bluetooth beacons. A kNN algorithm was used to calculate the distance between the current feature vector and collected training data. Based on the resulting distances, the person is assigned to one of the four rooms. Afterwards, a filter function is used to reduce false classifications. The idea is, that quick room hops are impossible and consequently room changes are only valid, if at least two consecutive samples have the same room label. Due to the low and non-constant sampling rate, there is no information about the location of a person between two scans in case of a room change. Therefore, the scan period was divided into two equal parts and it was assumed that the person stayed in both areas for the same duration. Finally, one room is assigned to each TSS (the one in which the person spent the most time during the TSS) and all objects that are not located within this room, but are included in the list of possible objects for the TSS are removed (see Figure 4.21).

Evaluation: The impact of different kNN k values on the average *recall* and *precision* was analyzed. Figure 4.22 visualizes achieved results. The following k values were considered: 1, 3, 5, 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60 and 70.

It can be seen that the result is quite robust for different k values. Nevertheless, the highest *recall* of 82% could be achieved with $k = 10$. The corresponding *precision* is 31%. This means, that when combining $BS_{optDistOH}$ with information about the persons' location at a room level, the *recall* of $BS_{optDistOH}$ could be raised by 7% to 82% and the related *precision* by 9% to 31%. Additionally, the EER could be improved by 9% to 56% (see Figure 4.23).

A deeper analysis of $recall_{obj}$, $precision_{obj}$ and EER_{obj} for each object shows that the average EER_{obj} could be raised significantly by 15% to 54% (see Table 4.9). The average $recall_{obj}$ and corresponding $precision_{obj}$ could even be raised by 8% and 9%. Beside these improvements, there was only one significant $precision_{obj}$ deterioration. The $precision_{obj}$ of

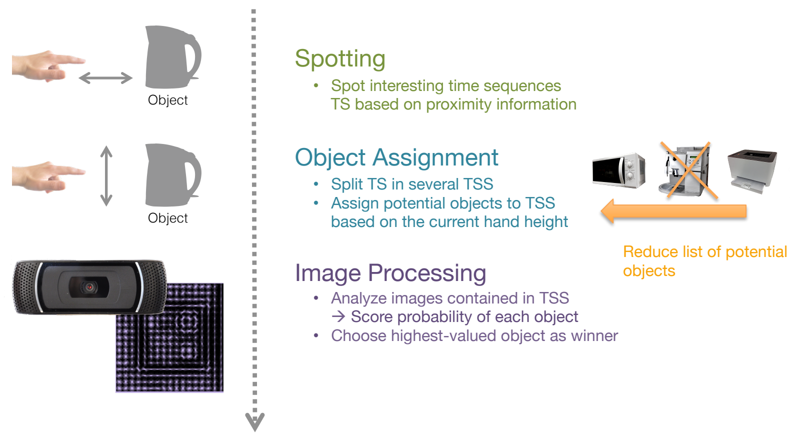


Figure 4.21: BS_{BT}: Integration Procedure

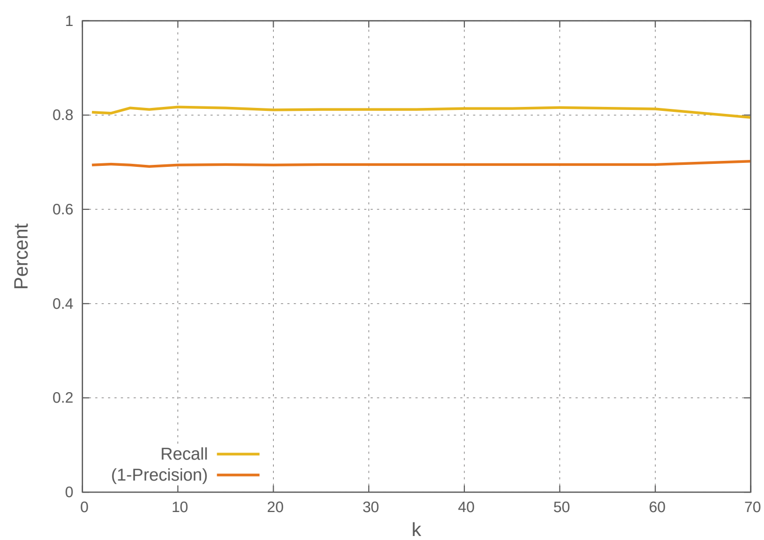


Figure 4.22: Analyzing impact on *recall* and (1−*precision*): k values of kNN algorithm.

"Air Conditioner" decreased by 37%, which means that the system was not able to provide the correct location at all times.

Besides *recall*, *precision* and *EER*, the effect on the amount of analyzed images and classification steps performed was evaluated. Table 4.10 compares BS_{BT} to previously introduced systems and shows the reduction of images and classification steps in percent. It can be seen that the amount of classification steps was reduced by significant 61%. This fact implies an immense improvement in terms of processing time. Apart from that, the number of analyzed images could be reduced by 4%.

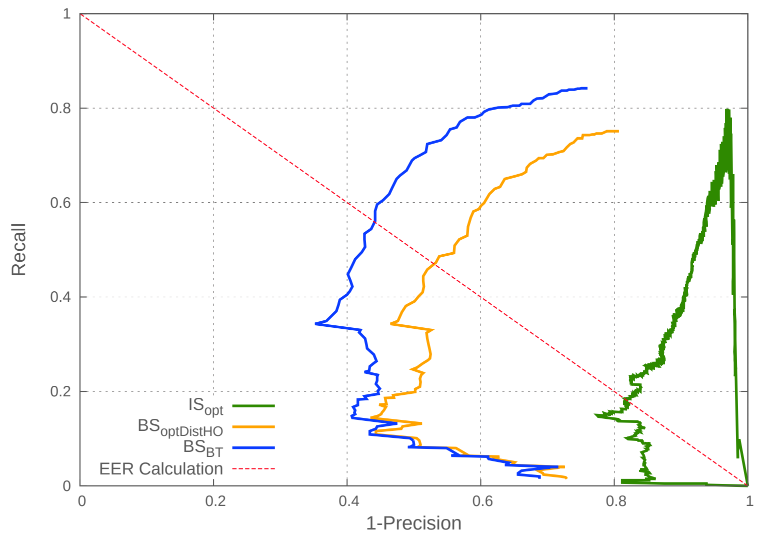


Figure 4.23: $(1-precision)-recall$ curves (based on $thr_{svmScore}$): BS_{BT} (blue curve; optimized basic system+room-level location), $BS_{optDistOH}$ (orange curve; optimized basic system) and IS_{opt} (green curve; inertial system).

Table 4.9: Object overview for BS_{BT}: $recall_{obj}$, $precision_{obj}$, EER_{obj} and improvements compared to $BS_{optDistOH}$ (optimized basic system) in terms of $recall_{obj}$ ($\triangle Rec$), $precision_{obj}$ ($\triangle Prec$) and EER_{obj} ($\triangle EER$).

Object	$recall_{obj}$	$precision_{obj}$	EER_{obj}	$\triangle Rec$	$\triangle Prec$	$\triangle EER$
Battery Charger	86	74	83	3	20	7
Coffee Machine	85	30	58	0	-1	2
PC	2	3	2	2	3	2
Air Conditioner	83	23	74	40	-37	74
Climatic Control Panel	82	54	70	0	28	29
Microwave	99	14	58	0	6	10
Ethernet Connector	73	21	59	0	11	6
Ring Binder	93	67	87	4	18	17
Power Socket	70	41	59	0	13	15
Laser Printer	89	15	46	26	13	27
Ink Printer	86	14	66	-3	9	19
Light-Shutter Switch	96	5	11	28	-2	-3
Scanner	98	10	25	3	3	0
Wall Cupboard	98	63	62	17	49	36
Cupboard	78	55	78	0	12	2
Water Tap	96	5	30	0	0	0
ØAverage	82	31	54	8	9	15

Table 4.10: System comparison – BS_{BT}: $recall$, $precision$, EER and percentage reduction of classification steps (CS_R) as well as analyzed images (AI_R) compared to $BS_{optDistOH}$ (optimized basic system).

System	$recall$	$precision$	EER	CS_R	AI_R
IS	80	3	18	–	–
$BS_{optDistOH}$	75	22	47	–	–
BS_{BT}	82	31	56	61	4

4.7.2.2 A Sub-Room Level Location System

System Setup: We have already seen that locating a person on a room level can cause an immense improvement on *recall*, *precision* and *EER*. This section introduces a sub-room level location system. It is clear that a system like this providing information such as "someone is currently located in front of the coffee machine" is much more powerful than just knowing that the person is currently located in the kitchen. As a low-cost and unobtrusive sensor system, ceiling mounted fish-eye cameras (Axis 212 PTZ; 30 frames per second at a resolution of 640x480 pixels; Price: about 580 €) were used. The big advantage of a fish-eye camera is, that a single device is able to cover a whole standard-sized room (see Figure 4.24).

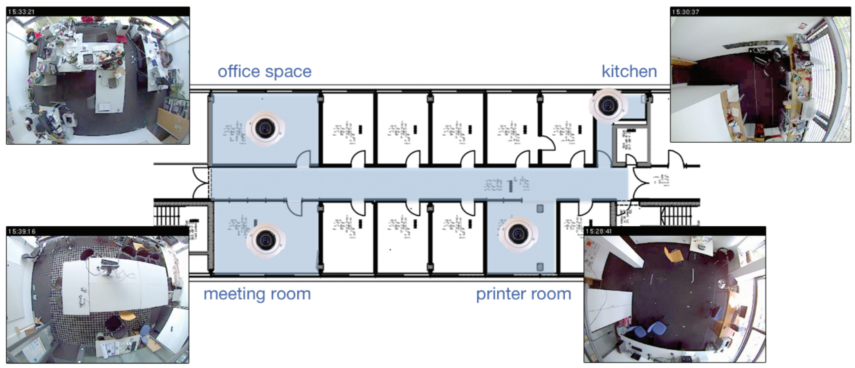


Figure 4.24: Floor Plan: Position of ceiling mounted fish-eye cameras.

On every image, polygonal lines mark objects or even a group of objects in case of nearby items. Chapter 2 already introduced a way of locating people at a sub-room level by combining information from ceiling-mounted cameras and motion sensors. However, a simpler approach based exclusively on cameras was used in this work. The system is able to detect general activities within pre-defined regions of interests (ROI) without identifying the person. It is worth noting, that the accuracy of the system may be significantly improved when using a more powerful sub-room level system like the one described in Chapter 2. The combination of $BS_{optDistOH}$ and the sub-room level location system is referred to as BS_{ROI}.

System Training and Configuration: The system setup includes one-time object/object group labellings on corresponding room images. Therefore, a simple program which captures images from Ethernet cameras and provides methods to define labeled polygonal lines (ROIs; see Figure 4.25 for labeled room images) was developed. Altogether 11 ROIs containing 16 objects (see Table 4.11) were defined.

Classification: The camera was configured to capture images with a resolution of 320x240 pixels and consecutive frames were used to calculate difference images. The idea is to distinguish between so-called background pixels (static background) and foreground pixels (movements/pixel changes). Therefore, a threshold $thrROI_{minPixDiff}$ on pixel difference values was used and pixels which have a higher difference value than the defined threshold are labeled as foreground pixels. Based on difference images, all foreground pixels within pre-defined ROIs are counted. A ROI is active (which means that a person is performing an activity within the ROI) if the amount of foreground pixels is higher than $thrROI_{minActPix}$. This way, smaller background changes, which are not caused by human activity, are rejected. Figure 4.26 shows examples of active ROIs and detected foreground pixels. As each activity lasts an observable amount of time, several single ROI activity events must be present within a TSS. Consequently, a ROI is active within a TSS only if the amount of related ROI activities is larger than $thrROI_{minTime}$.

Finally, objects belonging to non-active ROIs are removed from the list of possible TSS objects (see Figure 4.27). It is worth noting, that all experiments were performed within a real

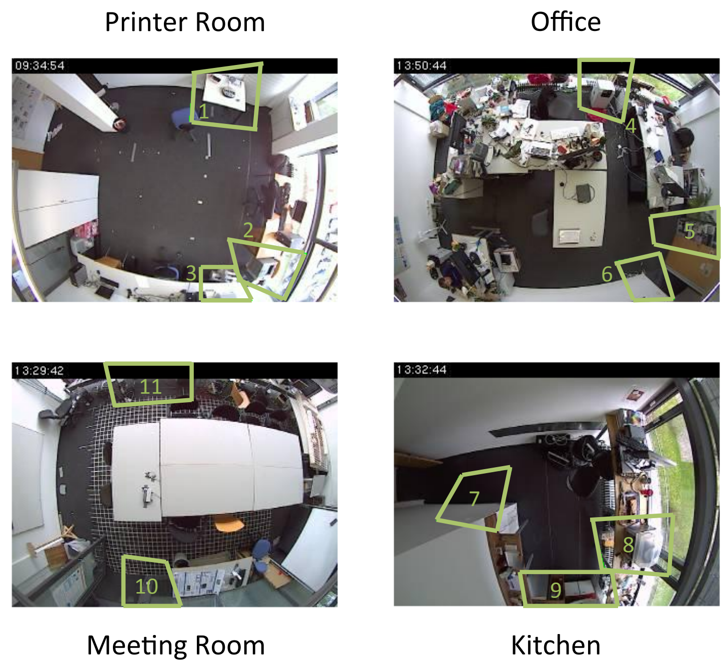


Figure 4.25: Labeled ROIs

Table 4.11: Object – ROI Assignment

ROI	Contained Objects
1	Battery Charger
2	Laser Printer
3	Ink Printer
4	Air Conditioner
5	Ring Binder
6	Light-Shutter Switch
7	Ethernet Connector
8	Microwave
9	Water Tap, Coffee Machine, Wall Cupboard and Cupboard
10	Climatic Control Panel
11	Scanner, PC

office environment during normal working hours. This means, that several people other than the experiment participants were moving or acting within pre-defined ROIs. This fact may lead to false detections as only anonymized information about ROI activity events were considered. The system described in Chapter 2 can be used to overcome this problem. Consequently, the following results can be improved by combining sub-room level location and person identification systems. Due to the applied processing steps, the system is able to handle about nine frames per second on a standard personal computer.

Evaluation: The impact of system parameters ($thrROI_{minTime}$, $thrROI_{minPixDiff}$, $thrROI_{minActPix}$) on *recall*, *precision* and EER was evaluated. Figure 4.28 visualizes the results. By setting parameters to (4, 1750, 50) the best *recall* of 90% with a corresponding *precision* of 40% is reached. As could be expected, a sub-room level location surveillance

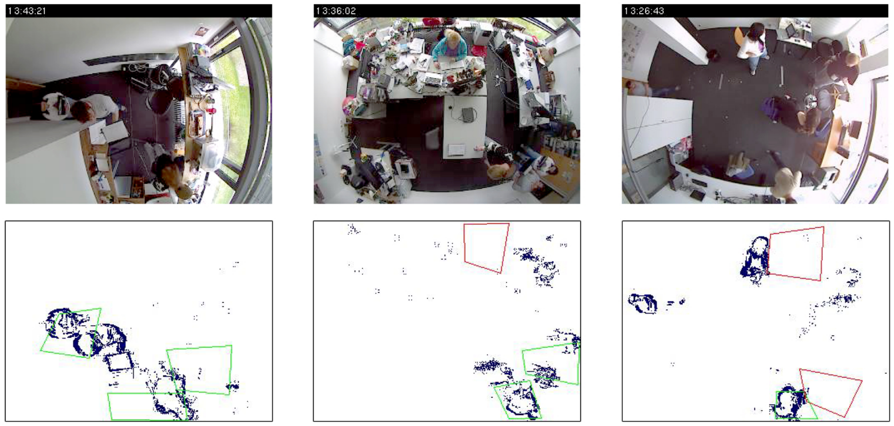


Figure 4.26: Images showing pre-defined (red rectangles) and active (green rectangles) ROIs based on detected foreground pixels (black pixels).

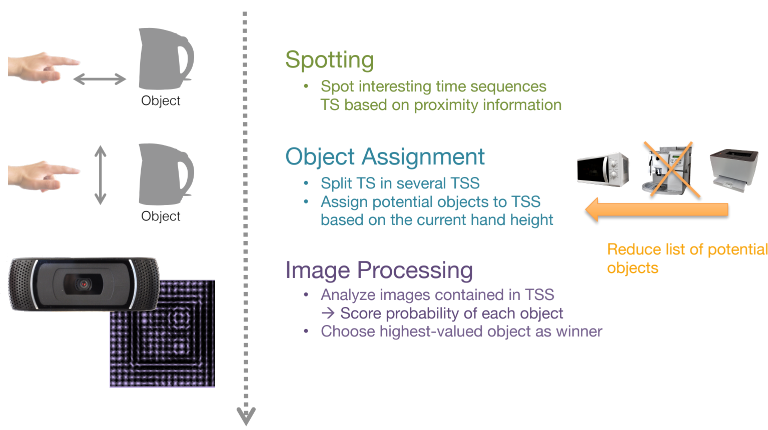


Figure 4.27: BS_{ROI}: Integration Procedure

system has much more impact on the recognition quality than systems providing only room level information. Additionally, the setup is much more easily than compared to the room level location system introduced as there is no need to collect reference data and the user is not forced to carry any additional devices (e.g. a smartphone). Compared to the room-level location system, the proposed approach is able to raise the *recall* by 8%, the *precision* by 9% and the EER by 5%. Figure 4.29 shows the $(1-precision)-recall$ curves resulting from various $thr_{svmScore}$ values. Compared to $BS_{optDistOH}$ a *recall* improvement of 15% and a *precision* improvement of 18% is achieved. The EER could be raised by 14% to 61%.

Table 4.12 shows $recall_{obj}$, $precision_{obj}$ and EER_{obj} for each specific object. It can be seen that for all objects except the "Cupboard" a $recall_{obj}$ improvement and for 12 objects a $precision_{obj}$ improvement (partly significant) was achieved. When looking at EER_{obj}, we can see that 13 objects improved (partly immense), two objects could retain their EER_{obj} ("PC", "Water Tap") and one object ("Coffee Machine") had a slight EER_{obj} decrease of 4%.

When analyzing the number of reduced images and classification steps, it can be seen that both values decreased significantly. Compared to $BS_{optDistOH}$ the amount of images analyzed was reduced by 28% and the amount of classifications even by 80%. As was already expected, BS_{ROI} is able to improve *recall*, *precision*, EER as well as reduce the amount of analyzed images and classification steps compared to BS_{BT} (see Table 4.13).

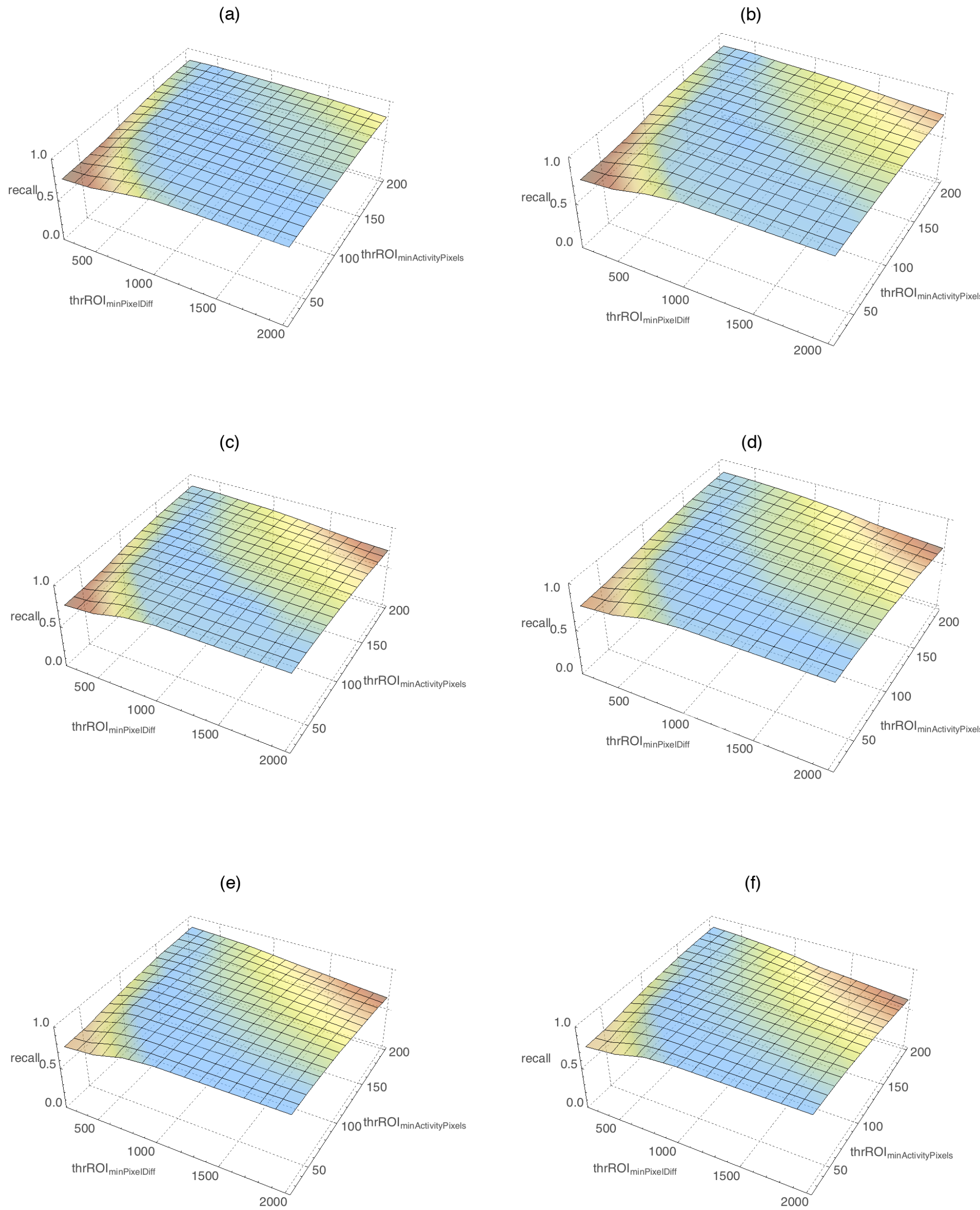


Figure 4.28: BS_{ROI}: Parameter Optimization. $thrROI_{minTime}$ ranges from 1 to 6 in steps of 1 (referenced as (a) - (f)); $thrROI_{minPixDiff}$ ranges from 250 to 2000 in steps of 250; $thrROI_{minActPix}$ ranges from 25 to 200 in steps of 25. The highest *recall* of 90% is reached with $thrROI_{minTime} = 4$, $thrROI_{minPixDiff} = 1750$ and $thrROI_{minActPix} = 50$.

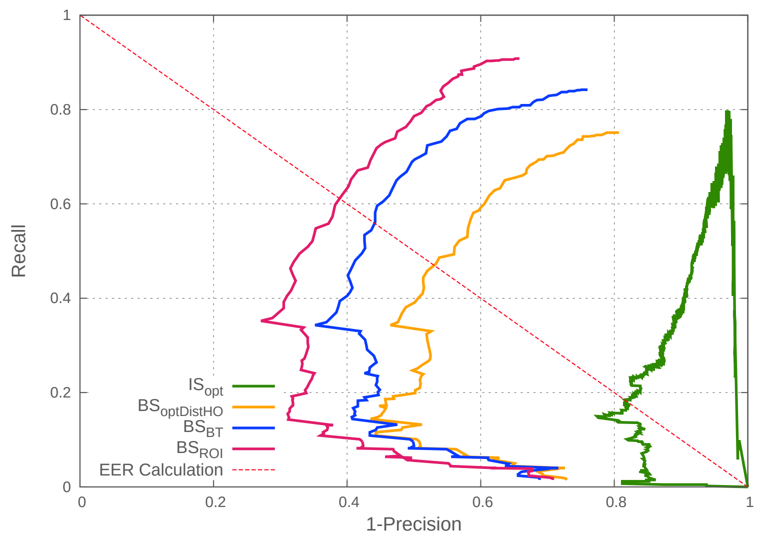


Figure 4.29: $(1-precision)-recall$ curves (based on $thr_{svmScore}$): IS_{opt} (green curve; inertial system), $BS_{optDistHO}$ (orange curve; optimized basic system), BS_{BT} (blue curve; optimized basic system+room-level location) and BS_{ROI} (pink curve; optimized basic system+sub-room level location).

Table 4.12: Object overview for BS_{ROI}: $recall_{obj}$, $precision_{obj}$, EER_{obj} and improvements compared to $BS_{optDistOH}$ (optimized basic system) in terms of $recall_{obj}$ ($\triangle Rec$), $precision_{obj}$ ($\triangle Prec$) and EER_{obj} ($\triangle EER$).

Object	$recall_{obj}$	$precision_{obj}$	EER_{obj}	$\triangle Rec$	$\triangle Prec$	$\triangle EER$
Battery Charger	93	48	90	10	-6	14
Coffee Machine	93	20	52	8	-11	-4
PC	22	41	0	22	41	0
Air Conditioner	87	63	78	44	3	78
Climatic Control Panel	96	76	91	14	50	50
Microwave	100	24	66	1	16	18
Ethernet Connector	90	21	60	17	11	7
Ring Binder	93	94	93	4	45	23
Power Socket	93	14	56	23	-14	12
Laser Printer	96	36	69	33	34	50
Ink Printer	96	47	78	7	42	31
Light-Shutter Switch	98	41	66	30	34	52
Scanner	100	19	30	5	12	5
Wall Cupboard	98	41	45	17	27	19
Cupboard	78	49	78	0	6	2
Water Tap	100	4	30	4	-1	0
ØAverage	90	40	61	15	18	22

Table 4.13: System comparison – BS_{ROI}: $recall$, $precision$, EER and percentage reduction of classification steps (CS_R) as well as analyzed images (AI_R) compared to $BS_{optDistOH}$ (optimized basic system).

System	$recall$	$precision$	EER	CS_R	AI_R
IS	80	3	18	–	–
$BS_{optDistOH}$	75	22	47	–	–
BS_{BT}	82	31	56	61	4
BS_{ROI}	90	40	61	80	28

4.7.2.3 Magnetic Field Signatures – Right Forearm Location

So far we have analyzed the impact of combining the basic system BS with a room level location and a sub-room level location system. Both systems need additional infrastructure equipment such as cameras or Bluetooth beacons. This section goes a step further and introduces a system which does not need any infrastructure and is used to determine the location/heading of the user's forearm based on magnetic field measurements (hereafter called BS_{RFL}).

System Setup: The system is based on a sensor mounted on a forearm, which is able to measure the magnetic field. In this work, the Xsens acceleration sensor, which is already included in the basic system BS, was used. Besides acceleration and Euler angles, the sensor provides a 3-dimensional magnetic field vector $v_{mag} = (mag_x, mag_y, mag_z)$.

System Training and Configuration: The system only needs a small set of magnetic field reference data for each object. Therefore, every fifth training sample was taken out of the pre-recorded Xsens data set (see Section 4.3.1). This corresponds to a sampling rate of 5 Hz. The reference data set includes magnetic field vectors describing the rough surrounding of each object. In real applications such data can be recoded easily during an initial setup phase where the person has to point at each object and its close surroundings once.

Classification: The idea behind this approach is, that a person must touch the object while interacting with it. Consequently, the user's hand is located close to the object and he/she has to point towards the object for a specific amount of time. During this time, the object specific magnetic field disturbance can be measured. Hence, the average magnetic field vector v_{avgMag} is calculated every half second and the closest Euclidean distance to each object is determined using corresponding reference data. For each TSS, the duration in which the distance is smaller than $thrMag_{minDist}$ is measured. If the duration for a specific object within a TSS is less than a pre-defined object independent threshold $thrMag_{minObjTime}$, it is removed from the list of possible TSS objects (see Figure 4.30).

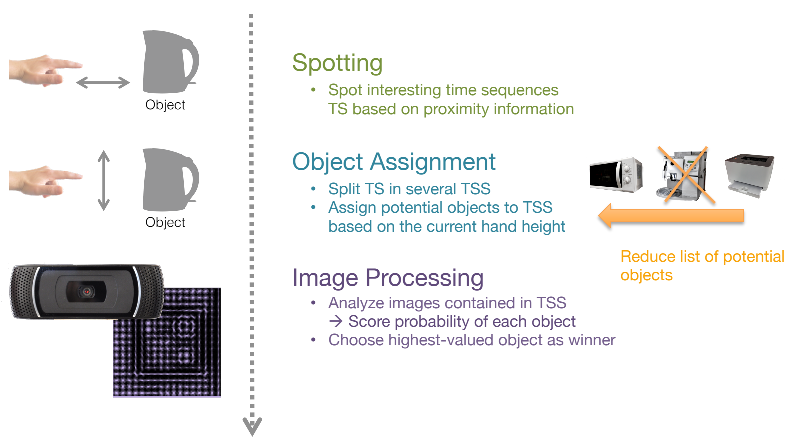


Figure 4.30: BS_{RFL}: Integration Procedure

Evaluation: The influence of $thrMag_{minDist}$ and $thrMag_{minObjTime}$ values on the system's recognition rate was analyzed. Therefore, $thrMag_{minDist}$ was varied from 0.1 mGauss to 1.0 mGauss in steps of 0.1 mGauss and $thrMag_{minObjTime}$ from 0.4 seconds to 2.9 seconds in steps of 0.1 seconds. Figure 4.31 visualizes the impact on the system's *recall*. The highest *recall* of 78% with a corresponding *precision* of 29% was achieved with $thrMag_{minDist} = 0.2$ mGauss and $thrMag_{minObjTime} = 0.6$ seconds. By fixing both parameters and varying $thr_{svmScore}$ a

EER of 55% could be achieved. Figure 4.32 shows the corresponding $(1 - precision) - recall$ curves.

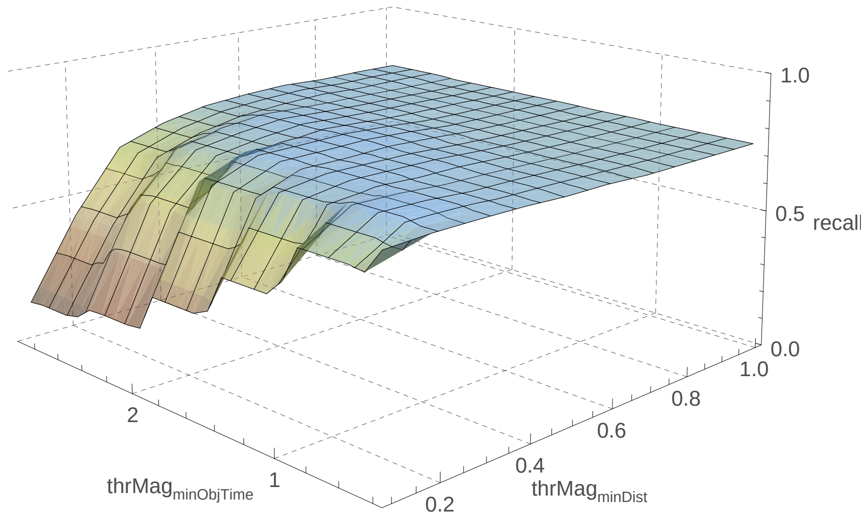


Figure 4.31: BS_{RFL}: Impact of $thrMag_{minObjTime}$ (unit: seconds) and $thrMag_{minDist}$ (unit: mGauss) on the system's *recall*.

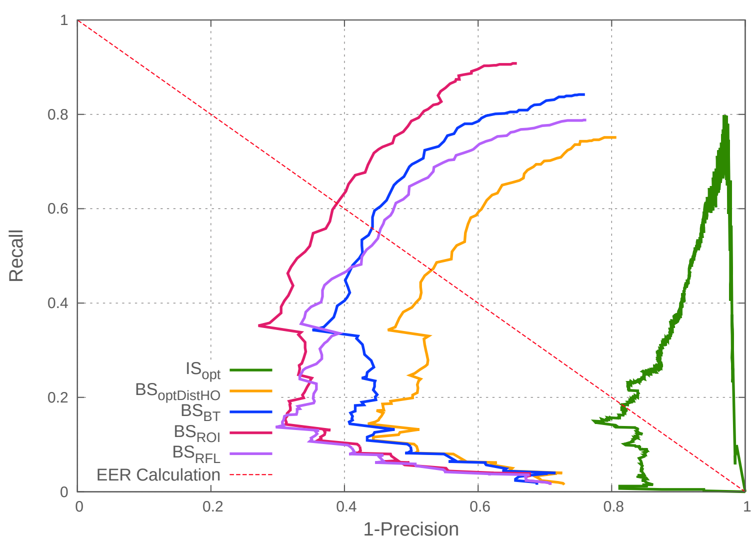


Figure 4.32: $(1 - precision) - recall$ curves (based on $thr_{svmScore}$): IS_{opt} (green curve; inertial system), $BS_{optDistHO}$ (orange curve; optimized basic system), BS_{BT} (blue curve; optimized basic system+room-level location), BS_{ROI} (pink curve; optimized basic system+sub-room level location) and BS_{RFL} (purple curve; optimized basic system+forearm location).

It can be seen, that this location method doesn't deliver as good results as BS_{BT} and BS_{ROI}. Nevertheless it is able to significantly improve the *recall* and *precision* of $BS_{optDistHO}$ without using any additional infrastructure. Table 4.14 shows $recall_{obj}$, $precision_{obj}$ and EER_{obj} as well as the improvements achieved compared to $BS_{optDistHO}$ for each object.

Table 4.14: Object overview for BS_{RFL}: $recall_{obj}$, $precision_{obj}$, EER_{obj} and improvements compared to $BS_{optDistOH}$ (optimized basic system) in terms of $recall_{obj}$ ($\triangle Rec$), $precision_{obj}$ ($\triangle Prec$) and EER_{obj} ($\triangle EER$).

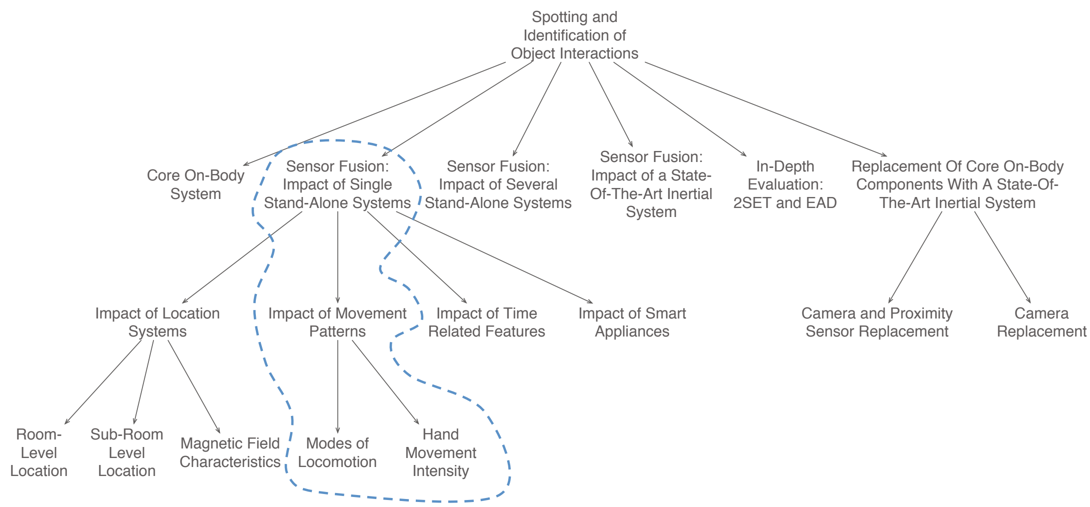
Object	$recall_{obj}$	$precision_{obj}$	EER_{obj}	$\triangle Rec$	$\triangle Prec$	$\triangle EER$
Battery Charger	90	19	76	7	-35	0
Coffee Machine	85	31	61	0	0	5
PC	4	1	0	4	1	0
Air Conditioner	65	59	64	22	-1	64
Climatic Control Panel	78	25	44	-4	-1	3
Microwave	99	12	56	0	4	8
Ethernet Connector	77	30	60	4	20	7
Ring Binder	93	72	87	4	23	17
Power Socket	44	54	0	-26	26	-44
Laser Printer	67	4	26	4	2	7
Ink Printer	96	6	50	7	1	3
Light-Shutter Switch	82	29	56	14	22	42
Scanner	95	9	30	0	2	5
Wall Cupboard	98	43	61	17	29	35
Cupboard	78	65	78	0	22	2
Water Tap	93	9	30	-3	4	0
ØAverage	78	29	49	3	7	10

The EER_{obj} of "Air Conditioner", "Ring Binder", "Light-Shutter Switch" and "Wall Cupboard" increased strongly. Only the EER of "Power Socket" decreased by 44%. All in all an average EER_{obj} improvement of 10%, an average $recall_{obj}$ improvement of 3% and an average $precision_{obj}$ improvement of 7% was achieved. Besides, a significant reduction of 19% with respect to the amount of images analyzed and of 66% in the case of classification steps performed could be reached (see Table 4.15).

Table 4.15: System comparison – BS_{RFL}: $recall$, $precision$, EER and percentage reduction of classification steps (CS_R) as well as analyzed images (AI_R) compared to $BS_{optDistOH}$ (optimized basic system).

System	$recall$	$precision$	EER	CS_R	AI_R
IS	80	3	18	–	–
$BS_{optDistOH}$	75	22	47	–	–
BS_{BT}	82	31	56	61	4
BS_{ROI}	90	40	61	80	28
BS_{RFL}	78	29	55	66	19

4.7.3 Sensor Fusion – Impact of Single Stand-Alone Systems: Movement Patterns



So far the influence of various location systems on the recognition quality of the problem addressed was analyzed. Besides location, information about general movements of people can be helpful to improve $BS_{optDistHO}$ as well. In the following, the focus will be on two systems, that are able to detect a person's mode of locomotion as well as the intensity of hand movements.

4.7.3.1 Modes of Locomotion

Almost all objects, that were considered in this work, are stationary objects (e.g. coffee machine, printer or cupboards). The idea is, that people are normally not interacting with such objects while they are walking. Consequently, images taken during walking periods were not analyzed. In the following, the combination of the basic system and modes of locomotion is referred to as BS_{MoL}.

System Setup: Various sensor modalities can be used to track people and to derive walking – standing activities. Examples are cameras (see [BL08] [YCKV07]), smart carpets (see [SL07] [LST13]) or acceleration sensors (see [TMN+12] [VLC00]). Based on such systems a person's current mode of locomotion can be determined. This work uses an acceleration sensor, which is already included in many mainstream smartphones to detect standing phases. During the experiments the smartphone was carried in the user's pocket. The built-in acceleration sensor delivers a three-dimensional acceleration signal (unit: m/s^2) $\overrightarrow{acc} = (acc_x, acc_y, acc_z)$ with a sampling rate of about 50 Hz.

System Training and Configuration: There are many different ways to recognize modes of locomotion using acceleration sensors. Depending on locomotion types in question, simple threshold based algorithms (see [BL08]) or more complex approaches (see [KNM+06] [BKV+97] [MCLC04]) are used to process raw acceleration data. This work aims to distinguish between "Standing" and "Walking" phases. As this problem is relatively simple and as it is assumed that the mobile remains in the pocket, there is no need to collect user dependent training data. Consequently, a threshold based algorithm is sufficient to solve the problem.

Classification: The Euclidean norm $|\overrightarrow{acc}|$ is calculated for each $\overrightarrow{acc}$. Using a fixed window length of 1 second, the variance over $|\overrightarrow{acc}|$ was determined. If the variance value was higher than $thrMoL_{var}$ the person is considered to be walking – otherwise we assume the person is stationary. As already mentioned, the idea is, that a person is not performing any object interaction during walking phases. Consequently images taken during that time were not processed (see Figure 4.33).

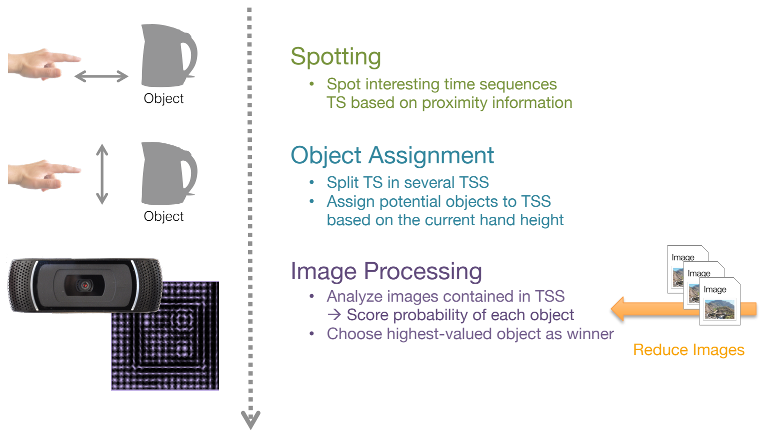


Figure 4.33: BS_{MoL}: Integration Procedure

Evaluation: Figure 4.34 shows the impact of several $thrMoL_{var}$ values ranging from 0.25 m^2/s^4 to 15 m^2/s^4 in steps of 0.25 m^2/s^4 on *recall* and *precision*. Both values are pretty stable having a $thrMoL_{var}$ value above 3 m^2/s^4. The highest *recall* of 75% with a corresponding *precision* of 22% is reached with $thrMoL_{var} = 8.5$ m^2/s^4. As is shown in Figure 4.35 the EER of BS_{MoL} is 47%. Ergo $thrMoL_{var}$ is not able to improve results achieved by $BS_{optDistHO}$. When taking a detailed look at specific objects this fact can be confirmed (see Table 4.16). No significant improvements were reached using this approach.

Nevertheless, the number of analyzed images as well as classification steps performed can be reduced by 3% each (see Table 4.17). In summary, it can be seen, that modes of locomotion have almost no influence on $BS_{optDistOH}$. The reason for this could be that $BS_{optDistOH}$ only takes intervals into account in which the person is close to an object and in most of the cases the person is already standing during such time periods.

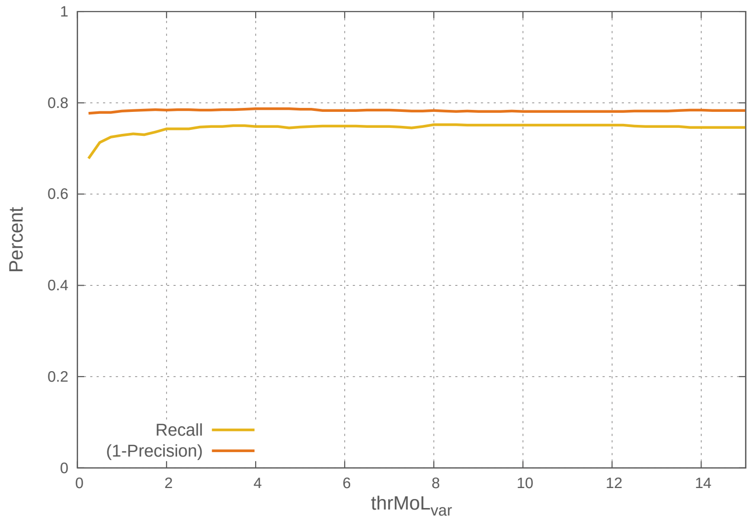


Figure 4.34: BS_{MoL}: *recall* and $(1-precision)$ depending on several $thrMoL_{var}$ values (unit: m^2/s^4).

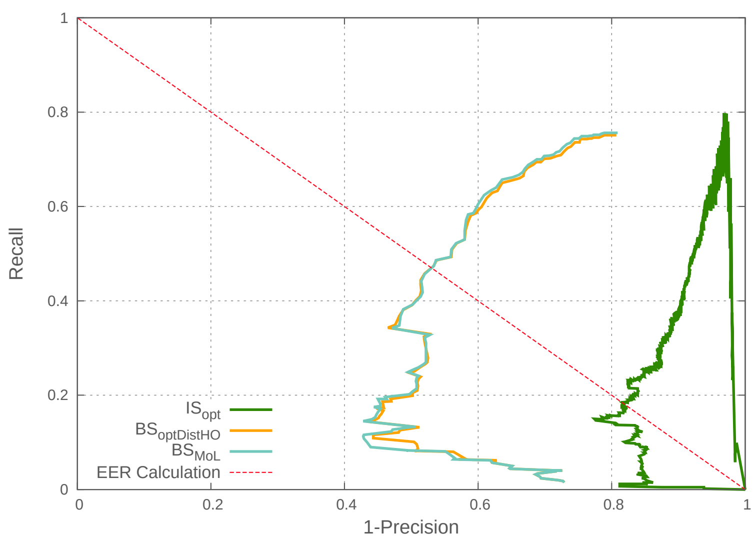


Figure 4.35: $(1-precision)-recall$ curves (based on $thr_{svmScore}$): IS_{opt} (green curve; inertial system), $BS_{optDistHO}$ (orange curve; optimized basic system) and BS_{MoL} (turquoise curve; optimized basic system+modes of locomotion).

Table 4.16: Object overview for BS_{MoL}: $recall_{obj}$, $precision_{obj}$, EER_{obj} and improvements compared to $BS_{optDistOH}$ (optimized basic system) in terms of $recall_{obj}$ ($\triangle Rec$), $precision_{obj}$ ($\triangle Prec$) and EER_{obj} ($\triangle EER$).

Object	$recall_{obj}$	$precision_{obj}$	EER_{obj}	$\triangle Rec$	$\triangle Prec$	$\triangle EER$
Battery Charger	83	54	76	0	0	0
Coffee Machine	85	31	59	0	0	3
PC	0	0	0	0	0	0
Air Conditioner	43	56	0	0	-4	0
Climatic Control Panel	82	26	44	0	0	3
Microwave	99	8	49	0	0	1
Ethernet Connector	73	10	53	0	0	0
Ring Binder	89	50	71	0	1	1
Power Socket	70	28	44	0	0	0
Laser Printer	63	2	15	0	0	-4
Ink Printer	89	5	47	0	0	0
Light-Shutter Switch	75	7	14	7	0	0
Scanner	95	7	25	0	0	0
Wall Cupboard	83	15	26	2	1	0
Cupboard	78	43	76	0	0	0
Water Tap	96	5	30	0	0	0
ØAverage	75	22	39	1	0	0

Table 4.17: System comparison – BS_{MoL}: $recall$, $precision$, EER and percentage reduction of classification steps (CS_R) as well as analyzed images (AI_R) compared to $BS_{optDistOH}$ (optimized basic system).

System	$recall$	$precision$	EER	CS_R	AI_R
IS	80	3	18	–	–
$BS_{optDistOH}$	75	22	47	–	–
BS_{MoL}	75	22	47	3	3

4.7.3.2 Hand Movement Intensity

Relevant activities and related object interactions emerge from subtle and similar hand movements. This section investigates, if the analysis of the intensity of the hand movement can help to improve the recognition rate of $BS_{optDistOH}$. The idea is that just before the person performs an activity, the hand moves quite fast in order to touch the object. Hence, time intervals showing almost no movement can be neglected. In the following, the combination of the basic system and intensity of hand movement is referred to as BS_{HM}.

System Setup: In order to calculate the intensity of hand movement a acceleration sensor mounted on the forearm was used. Such a sensor is already included in $BS_{optDistOH}$ and hence there is no need to add additional sensors or infrastructure.

System Training and Configuration: Recognizing arm movement intensity based on acceleration sensors is quite a simple task. For this purpose, a recognition system can be deployed without any user dependent training data or configuration.

Classification: A fixed window length of 13 samples (about 0.5 seconds) was used. For each window, the variance value using the Euclidean norm $|\overrightarrow{acc}|$ of the included samples was calculated. If the variance value was lower than $thrHM_{varThr}$, no significant hand movement was expected during that time. Hence, all images taken during this time were disregarded (see Figure 4.36).

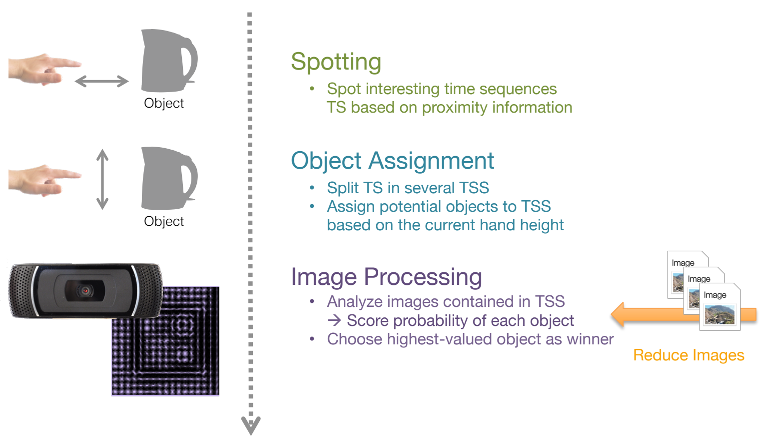


Figure 4.36: BS_{HM}: Integration Procedure

Evaluation: Various values of $thrHM_{varThr}$ ranging from 0 m^2/s^4 to 3 m^2/s^4 in steps of 0.025 m^2/s^4 were evaluated. The impact on *recall* and *precision* is visualized in Figure 4.37. The *recall* decreases significantly for $thrHM_{varThr}$ values above 0.2 m^2/s^4 and the error increases slightly for values above 0.25 m^2/s^4. The highest *recall* of 75% could be achieved by $thrHM_{varThr} = 0.1$ m^2/s^4. The corresponding *precision* is 22%. Therefore, the combination of $BS_{optDistOH}$ and information about the hand movement intensity is not able to improve the recognition quality. On the contrary, the EER even decreases by 1% to 46% (see Figure 4.38).

Looking at specific objects this fact can be confirmed. As table 4.18 shows, neither the average $recall_{obj}$, the average $precision_{obj}$ nor the average EER_{obj} could be improved. Nevertheless, the number of analyzed images could be reduced by significant 42% and classification steps performed by significant 35% (see Table 4.19).

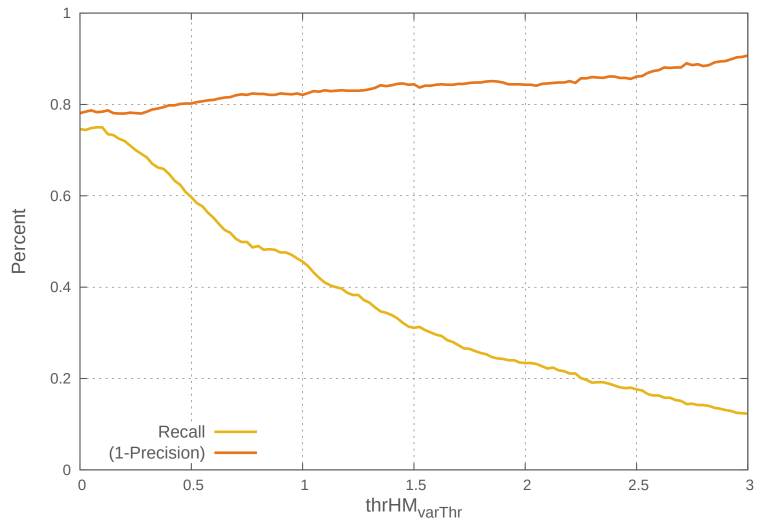


Figure 4.37: BS_{HM}: *recall* and (1−*precision*) in response to several $thrHM_{varThr}$ values (unit: m^2/s^4).

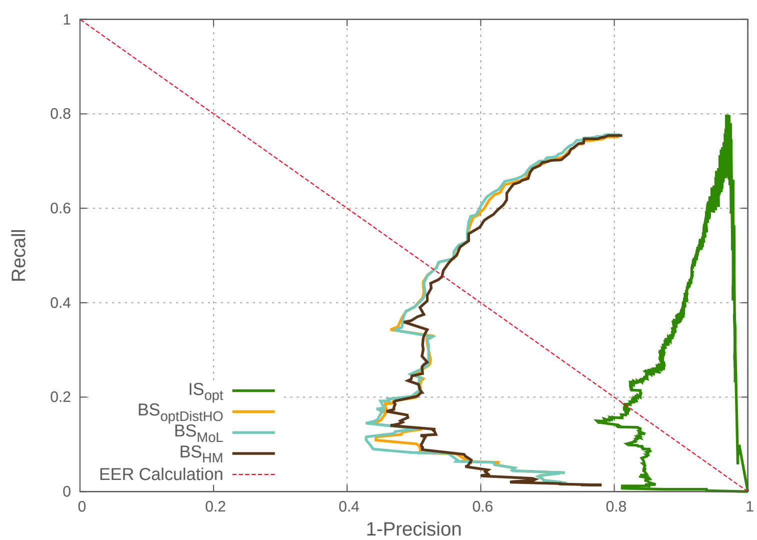


Figure 4.38: $(1 - precision) - recall$ curves (based on $thr_{svmScore}$): IS_{opt} (green curve; inertial system), $BS_{optDistHO}$ (orange curve; optimized basic system), BS_{MoL} (turquoise curve; optimized basic system+mode of locomotion) and BS_{HM} (dark brown curve; optimized basic system+hand movement).

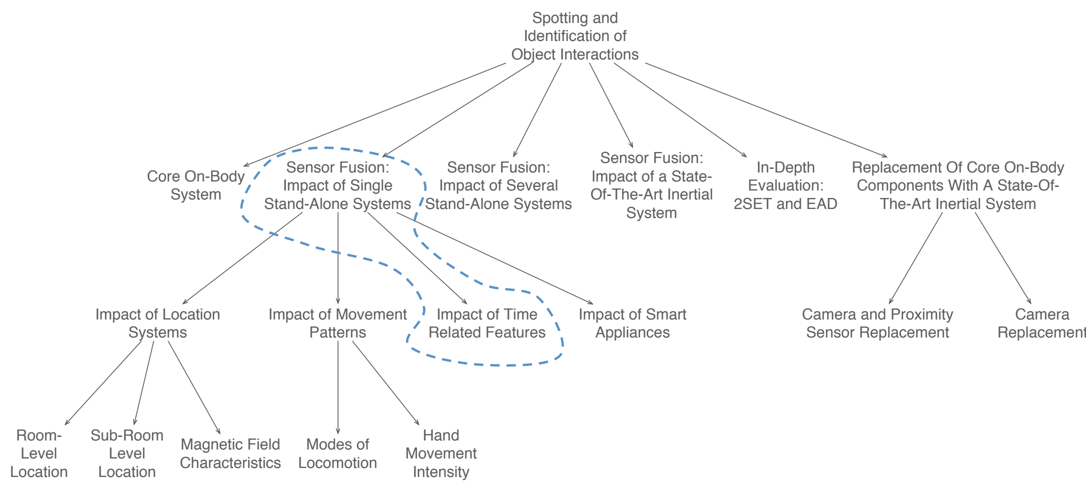
Table 4.18: Object overview for BS_{HM}: $recall_{obj}$, $precision_{obj}$, EER_{obj} and improvements compared to $BS_{optDistOH}$ (optimized basic system) in terms of $recall_{obj}$ ($\triangle Rec$), $precision_{obj}$ ($\triangle Prec$) and EER_{obj} ($\triangle EER$).

Object	$recall_{obj}$	$precision_{obj}$	EER_{obj}	$\triangle Rec$	$\triangle Prec$	$\triangle EER$
Battery Charger	76	49	72	-7	-5	-4
Coffee Machine	89	37	59	4	6	3
PC	0	0	0	0	0	0
Air Conditioner	46	52	0	3	-8	0
Climatic Control Panel	82	26	37	0	0	-4
Microwave	99	8	49	0	0	1
Ethernet Connector	83	11	60	10	1	7
Ring Binder	89	50	70	0	1	0
Power Socket	67	27	51	-3	-1	7
Laser Printer	63	2	20	0	0	1
Ink Printer	93	5	46	4	0	-1
Light-Shutter Switch	71	7	14	3	0	0
Scanner	94	7	23	-1	0	-2
Wall Cupboard	81	14	25	0	0	-1
Cupboard	78	43	76	0	0	0
Water Tap	89	5	30	-7	0	0
ØAverage	75	21	40	0	0	0

Table 4.19: System comparison – BS_{HM}: $recall$, $precision$, EER and percentage reduction of classification steps (CS_R) as well as analyzed images (AI_R) compared to $BS_{optDistOH}$ (optimized basic system).

System	$recall$	$precision$	EER	CS_R	AI_R
IS	80	3	18	–	–
$BS_{optDistOH}$	75	22	47	–	–
BS_{MoL}	75	22	47	3	3
BS_{HM}	75	22	46	35	42

4.7.4 Sensor Fusion – Impact of Single Stand-Alone Systems: Time Related Features



The basic system $BS_{optDistOH}$ uses proximity information to spot interesting time periods when people are close to objects. So far no time information is used to filter out spotted periods lasting an unusual amount of time compared to the standard duration for object interactions. For example, some people rest their arms against the wall while talking. As the distance to the wall is quite short during that time, $BS_{optDistOH}$ will spot a very long time interval as TS_i. Therefore, the idea of this section is to drop such intervals, which could lead to an improved recognition quality, reduced amount of classification steps and analyzed images. In the following the combination of the basic system and temporal features is referred to as BS_{TF}.

System Setup: Additional equipment is not needed.

System Training and Configuration: The system does not need any training data or extensive configuration. Only an overall time threshold based on the maximum activity duration must be set.

Classification: If the time duration of TSS_{i_j} lasts longer than a time threshold $thrTF_{maxDur}$, the TSS_{i_j} and included images were deleted (see Figure 4.39).

Evaluation: Several values for $thrTF_{maxDur}$ ranging from 1 to 15 seconds in steps of 1 second were evaluated. Figure 4.40 visualizes the impact on *recall* and *precision*. It can be seen, that $thrTF_{maxDur}$ values higher over six seconds have almost no effect on the system. It is obvious that object-specific time thresholds will result quite certainly in better system improvements. However, this fact is not considered in this work, as it implies an object related configuration step and would consequently destroy the idea of having an easy to train system. The highest *recall* of 75% with a corresponding *precision* of 22% was achieved by $thrTF_{maxDur} = 8$ seconds. The EER was 47%. Compared to $BS_{optDistOH}$, no improvements were achieved with respect to *recall*, *precision* and EER (see Figure 4.41). This fact was also confirmed when analyzing the

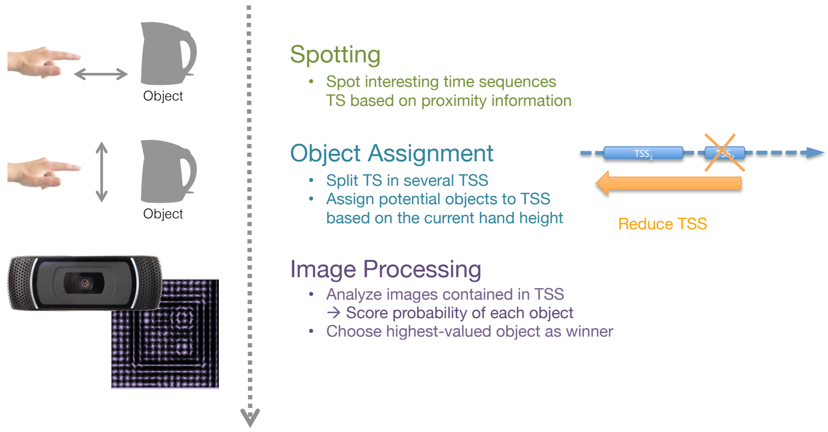


Figure 4.39: BS_{TF}: Integration Procedure

classification rate for specific objects (see Table 4.20). However, the number of analyzed images could be reduced by 21% and the number of classification steps by 14% (see Table 4.21).

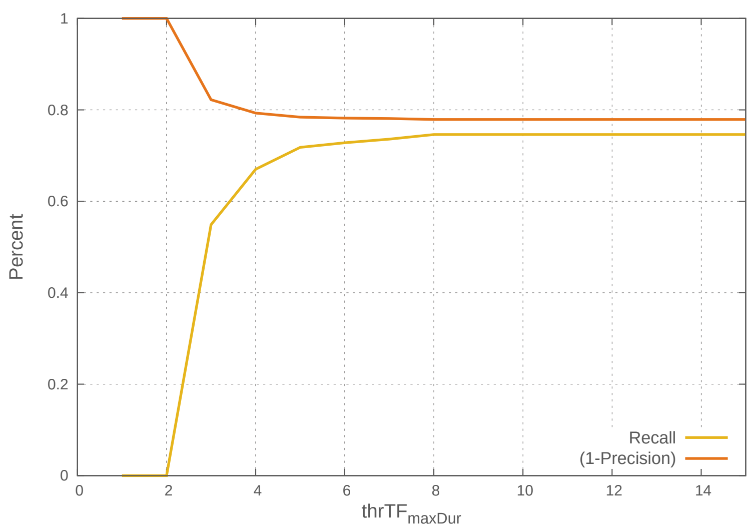


Figure 4.40: BS_{TF}: *recall* and $(1-precision)$ in response to $thrTF_{maxDur}$ values.

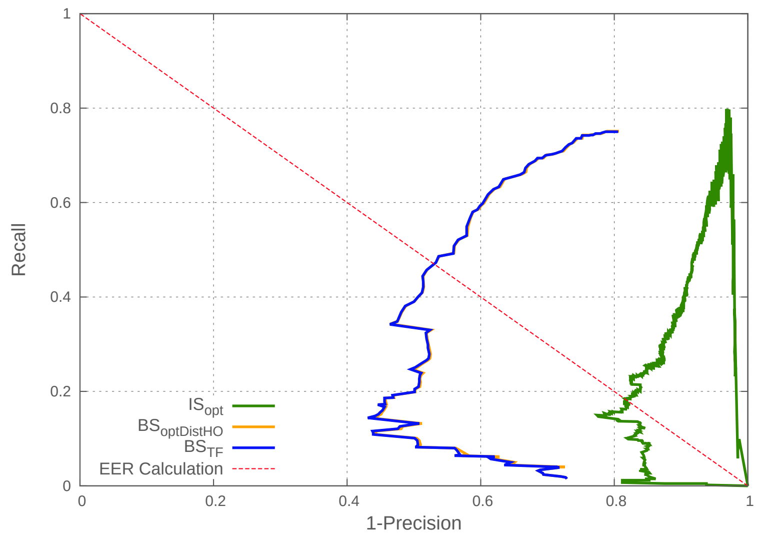


Figure 4.41: $(1-precision)-recall$ curves (based on $thr_{svmScore}$): IS_{opt} (green curve; inertial system), $BS_{optDistHO}$ (orange curve; optimized basic system) and BS_{TF} (blue curve; optimized basic system+temporal feature).

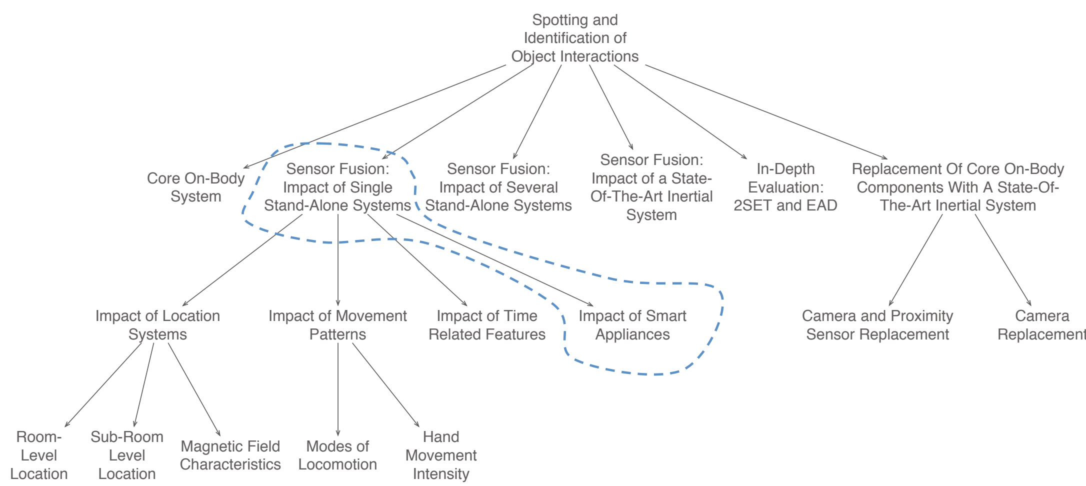
Table 4.20: Object overview for BS_{TF}: $recall_{obj}$, $precision_{obj}$, EER_{obj} and improvements compared to $BS_{optDistOH}$ (optimized basic system) in terms of $recall_{obj}$ ($\triangle Rec$), $precision_{obj}$ ($\triangle Prec$) and EER_{obj} ($\triangle EER$).

Object	$recall_{obj}$	$precision_{obj}$	EER_{obj}	$\triangle Rec$	$\triangle Prec$	$\triangle EER$
Battery Charger	83	56	76	0	2	0
Coffee Machine	85	31	56	0	0	0
PC	0	0	0	0	0	0
Air Conditioner	43	60	0	0	0	0
Climatic Control Panel	82	26	41	0	0	0
Microwave	99	8	48	0	0	0
Ethernet Connector	73	10	53	0	0	0
Ring Binder	89	49	70	0	0	0
Power Socket	70	28	44	0	0	0
Laser Printer	63	2	19	0	0	0
Ink Printer	89	5	47	0	0	0
Light-Shutter Switch	68	7	14	0	0	0
Scanner	94	8	25	-1	1	0
Wall Cupboard	81	14	26	0	0	0
Cupboard	78	43	76	0	0	0
Water Tap	96	5	30	0	0	0
ØAverage	75	22	39	0	0	0

Table 4.21: System comparison – BS_{TF}: $recall$, $precision$, EER and percentage reduction of classification steps (CS_R) as well as analyzed images (AI_R) compared to $BS_{optDistOH}$ (optimized basic system).

System	$recall$	$precision$	EER	CS_R	AI_R
IS	80	3	18	–	–
$BS_{optDistOH}$	75	22	47	–	–
BS_{TF}	75	22	47	14	21

4.7.5 Sensor Fusion – Impact of Single Stand-Alone Systems: Smart Appliances



During the last few years more and more high-end electronic devices such as printers, ovens or light switches turned into smart devices. These devices are able to provide information about their current operating modes (e.g. Miele@Home[51]). However, so far they have not found their way into common households as they are still too expensive and don't provide a standard interface. Chapter 3 introduced sensor systems able to turn mainstream household devices into smart devices. These systems were considered again in order to improve the recognition quality of $BS_{optDistOH}$.

4.7.5.1 System Setup

All in all six electronic devices and a water tap were turned into smart devices. In terms of electronic devices, a system comparable to the one introduced in Section 3.5 was used. Several appliances were attached to wireless power measurement sensors manufactured by the company "PLOGG International". Each device was connected to a single power sensor providing power characteristics such as power, phase angle, current and voltage every two seconds (see Figure 4.42). A single sensor device costs about 100 €. Pre-defined and device-specific rules based on power characteristics were used to recognize operating modes. In order to recognize the current state of the water tap as well as the approximate amount of flowing water, a similar system setup as described in Section 3.4 was chosen. An off-the-shelf Bluetooth headset microphone was fixed to a small box filled with insulation material. Afterwards, the box was mounted on the inflow pipe of the water tap. Sound samples were transmitted via Bluetooth to a central processing unit where the amount of water consumed was approximated. In the following, the combination of the basic system and smart appliances is referred to as BS_{SA}.

4.7.5.2 System Training and Configuration

Operating modes of electronic devices are defined by simple rules based on features related to power usage, reactive power values and time. As already mentioned, such rules can either be

[51] http://www.miele-at-home.de/de/aktion/mieleathome/656.htm (last accessed on 2013/06/19)

Figure 4.42: Left: PLOGG power sensor; Right: Plogg sensor connected to a coffee machine.

created by the user (with the help of simple software tools) or they can be provided by device manufacturers. In this scenario, the following devices and corresponding operating modes were monitored:

- Battery charger (insert empty battery, remove battery while charging)
- PC (turn on, turn off)
- Microwave (door open, door closed, heating)
- Coffee machine (make espresso, make coffee)
- Air conditioner (on, off)
- Scanner (on, off, scanning)

Algorithms 1-6 show defined rule sets for each device. It is worth noting, that all rules were adapted to the considered activity set and therefore rules may be different if additional activities should be also considered.

if *($Power_{cur} \geq 30.0W$)* **then**
    $counter_{on} \leftarrow counter_{on} + 1$
    if *($status_{cur} == off$)* **then**
        $status_{cur} = on$
        Event: PC on
else
    if *($status_{cur} == on$)* **then**
        if *($counter_{on} > 2$)* **then**
            Event: PC off
        $status_{cur} = off$
        $counter_{on} \leftarrow 0$

Algorithm 1: Rule set describing operating modes of PC

if *($Power_{cur} > 50.0W$)* **then**
  $counter_{on} \leftarrow counter_{on} + 1$
  if *($status_{cur} == off$)* **then**
    $status_{cur} \leftarrow on$
    Event: Air Conditioner on
else
  if *($status_{cur} == on$)* **then**
    if *($counter_{on} > 2$)* **then**
      Event: Air Conditioner off
    $counter_{on} \leftarrow 0$
    $status_{cur} \leftarrow off$

Algorithm 2: Rule set describing operating modes of Air Conditioner

if *($Power_{cur} > 8.0W$) $\wedge$ ($Power_{cur} < 25.0W$)* **then**
  $counter_{open} \leftarrow counter_{open} + 1$
  if *($status_{cur} == idle$)* **then**
    Event: Door open
    $status_{cur} = open$
else if *($Power_{cur} > 1000.0W$)* **then**
  if *($status_{cur} == heating$)* **then**
    Event: Heating
    $status_{cur} \leftarrow heating$
else
  if *($status_{cur} == open$) $\wedge$ ($counter_{open} > 1$)* **then**
    Event: Door Closed
  if *($status_{cur} == heating$)* **then**
    Event: Heating Finished
  $status_{cur} \leftarrow idle$
  $counter_{open} \leftarrow 0$

Algorithm 3: Rule set describing operating modes of Microwave

```
if (Power_cur ≤ 6.0W) then
    if (deviceOff == false) then
        if (status_cur == on) then
            status_cur ← off
            Event: Scanner off
    else
        deviceOff ← true
        if (status_cur == scanning) then
            status_cur ← on
            if (sizeOf(List_PowerValues) > 3) then
                calculate variance value var of List_PowerValues
                if (var < 0.1) then
                    Event: Scan finished
    checkScan ← false
    clear List_PowerValues
else if (Power_cur > 11.0W) ∧ (Power_cur < 12.0W) then
    if (checkScan == true) then
        deviceOff ← false
        List_PowerValues add Power_cur
        if (status_cur ≠ scanning) then
            status_cur ← scanning
    else
        checkScan ← true
else if (Power_cur > 6.0W) then
    deviceOff ← false
    checkScan ← false
    if (sizeOf(List_PowerValues) > 3) then
        calculate variance value var of List_PowerValues
        if (var < 0.1) then
            Event: Scan finished
    else
        checkScan ← false
    if (status_cur == scanning) ∨ (status_cur == off) then
        Event: Scanner on
        status_cur ← on
    clearList_PowerValues
else
    deviceOff ← false
    clearList_PowerValues
```

Algorithm 4: Rule set describing operating modes of Scanner

if *($ReactivePower_{cur} > -39.0var$)* **then**
    if *($Status_{cur} == makeCoffee$)* **then**
        if *($sizeOf(List_{ReactivePowerValues}) > 3$)* **then**
            if *($sizeOf(List_{ReactivePowerValues}) \leq 8$)* **then**
                Event: Espresso
            else
                Event: Coffee
        Clear $List_{ReactivePowerValues}$
        $Status_{cur} \leftarrow idle$
else
    $List_{ReactivePowerValues}$ add $ReactivePower_{cur}$
    $Status_{cur} \leftarrow makeCoffee$

Algorithm 5: Rule set describing operating modes of Coffee Machine

if *($Power_{cur} < 4.0W$)* **then**
    if *($Status_{cur} == charging$)* **then**
        $Status_{cur} \leftarrow idle$
        Event: Battery removed while charging
else
    if *($Status_{cur} == idle$)* **then**
        $Status_{cur} \leftarrow charging$
        Event: Empty battery inserted

Algorithm 6: Rule set describing operating modes of Battery Charger

The usage of power sensors is rather simple as they can be operated out-of-the-box, if device manufacturers provide the necessary rule sets. The configuration of the water measurement sensor system is almost as simple as turning electronic devices into smart devices. Once the water sensor is attached to the inflow pipe, the user has to record a small set of sound samples for each of the desired water flows as well as for silence. In order to be robust against background noise, a threshold value has to be adjusted to the environment. These steps can be performed within a short time using a simple software tool. In this work four levels of water flow are considered (see Table 4.22).

Table 4.22: Water measurement system: Water flows considered

Water Level	Water Flow (ml/sec)
0	–
1	83.33
2	177.77
3	228.57

4.7.5.3 Classification

The classification procedure is very straightforward. With respect to electronic devices, the power used and reactive power values are applied to pre-defined rule sets. The event of a changing operating mode is used to reduce the amount of possible objects included in each TSS_{i_j}. The idea is, that object interactions imply operating mode changes. If a specific electronic device has not modified its current operating mode during a TSS_{i_j}, it is assumed, that this device was not used and is consequently removed from the list of possible objects. Due to the fact, that devices need some seconds until they respond to the user's input and due to the design of the processing algorithms used, each TSS_{i_j} was virtually extended by +/- 3 seconds when comparing it with the operating mode changes detected[52].

Regarding the water flow measurement system, the classification procedure is almost the same as described in Chapter 3.4. In order to reduce background noise and to also reject false classifications the following conditions must be fulfilled:

- $kNN_{dist} < 3600$
- It is assumed, that no water is running if three consecutively analyzed sound samples indicate silence or background noise.

Figure 4.43 visualizes the integration procedure into $BS_{optDistOH}$ again.

4.7.5.4 Evaluation

Compared to $BS_{optDistOH}$, BS_{SA} was able to increase the *recall* by 6% to 81%. The corresponding *precision* was raised by 13% to 35% and the resulting *EER* was improved by 6% to 53%. Figure 4.44 shows the associated $(1 - precision) - recall$ curves.

When analyzing the impact of BS_{SA} on specific devices, it can be seen, that a significant precision improvement could be achieved for many devices (e.g. "Water Tap", "Coffee Machine" or "Microwave"). In almost all cases the recall was held or even improved (e.g. "Laser Printer" or "Ethernet Connector"). Some of the devices were not turned into smart devices. Nevertheless, it seems that false classifications were properly filtered and rated down, and therefore correctly classified objects were chosen as final winners. Only the "Coffee Machine" and the "Microwave" lost 11% and 1% of their previous $recall_{obj}$. This means, that the rules defined were not able

[52] In this work "Clean Microwave" events follow immediately after a "Door open" event and therefore the related TSS_{i_j} are covered by such events. Without this pre-condition, a different operating mode rule set for the microwave device has to be defined to integrate "cleaning" events.

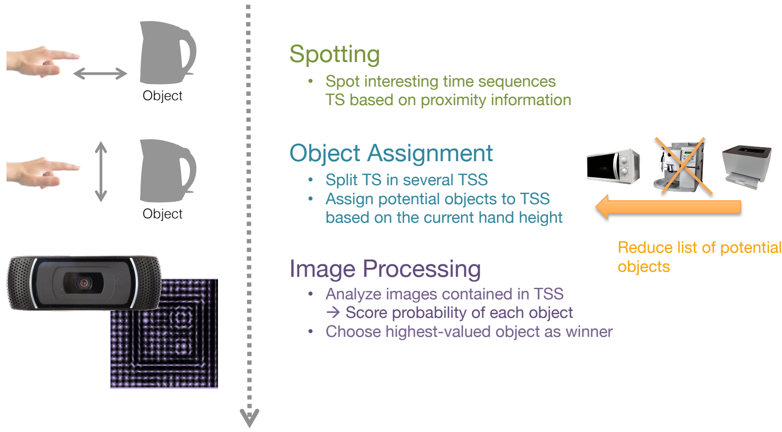


Figure 4.43: BS_{SA}: Integration Procedure

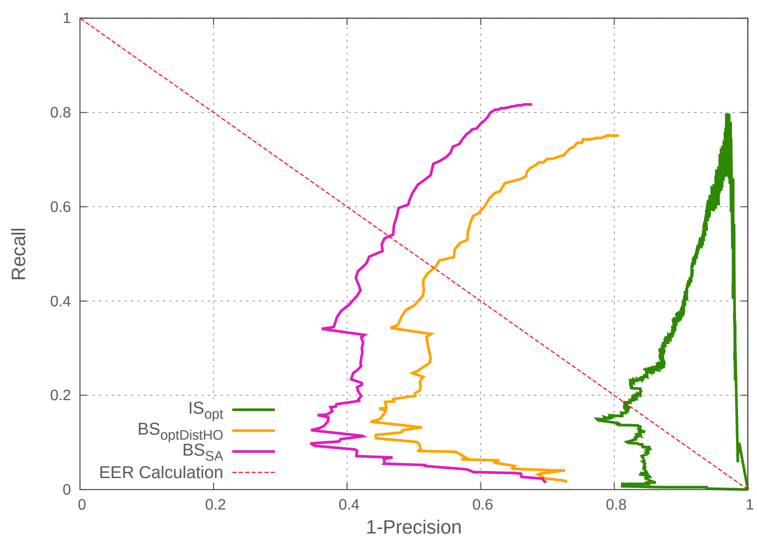


Figure 4.44: $(1 - precision) - recall$ curves (based on $thr_{svmScore}$): IS_{opt} (green curve; inertial system), $BS_{optDistHO}$ (orange curve; optimized basic system) and BS_{SA} (purple curve; optimized basic system+smart appliances).

to reliably describe events such as "make coffee". The EER_{obj} was increased significantly for "Microwave", "Scanner" and "Water Tap". Unfortunately, we can see a strong EER_{obj} decrease in the case of the "Coffee Machine". The EER for this device went down to zero. As this behavior is not reasonable, the EER_{obj} for this object is analyzed more specifically. Figure 4.45 shows again, that the precision of "Coffee Machine" significantly improved compared to $BS_{optDistOH}$. However, the highest $recall_{obj}$ achieved is still lower than the highest $recall_{obj}$ of $BS_{optDistOH}$ and from a mathematical standpoint the EER_{obj} is zero. All in all the average $recall_{obj}$ could be increased by 7%, the average $precision_{obj}$ by 13% and the average EER_{obj} by 3%. It is emphasized, that in case of the "Air Conditioner" a $precision_{obj}$ of 100% is reached while still having a $recall_{obj}$ of 65%. Besides, BS_{SA} is not able to reduce the number of analyzed images. However, the number of classification steps was reduced by 51% (see Table 4.24).

Smart devices were not exclusively used by experiment participants. As the data recording took place during normal working hours, many people apart from the experiment participants used the electronic devices and the water tap. All in all 26031 operating mode changes were registered. This exceptionally large number of operating mode changes could not have been caused exclusively by real device interactions. When analyzing the operating mode changes recognized for each device, it turned out that 21771 events were registered for the "Battery Charger". Consequently, the chosen rules for this device were not able to detect real charging events with a high precision. All in all 356 object interactions were performed by experiment participants.

Table 4.23: Object overview for BS_{SA}: $recall_{obj}$, $precision_{obj}$, EER_{obj} and improvements compared to $BS_{optDistOH}$ (optimized basic system) in terms of $recall_{obj}$ ($\triangle Rec$), $precision_{obj}$ ($\triangle Prec$) and EER_{obj} ($\triangle EER$).

Object	$recall_{obj}$	$precision_{obj}$	EER_{obj}	$\triangle Rec$	$\triangle Prec$	$\triangle EER$
Battery Charger	83	69	79	0	15	3
Coffee Machine	74	80	0	-11	49	-56
PC	0	0	0	0	0	0
Air Conditioner	65	100	0	22	40	0
Climatic Control Panel	85	19	39	3	-7	-2
Microwave	98	75	88	-1	67	40
Ethernet Connector	90	7	57	17	-3	4
Ring Binder	89	49	70	0	0	0
Power Socket	78	12	44	8	-16	0
Laser Printer	96	2	17	33	0	-2
Ink Printer	93	3	46	4	-2	-1
Light-Shutter Switch	96	6	11	28	-1	-3
Scanner	99	36	47	4	29	22
Wall Cupboard	81	14	26	0	0	0
Cupboard	78	38	76	0	-5	0
Water Tap	96	51	74	0	46	44
ØAverage	81	35	42	7	13	3

Table 4.24: System comparison – BS_{SA}: $recall$, $precision$, EER and percentage reduction of classification steps (CS_R) as well as analyzed images (AI_R) compared to $BS_{optDistOH}$ (optimized basic system).

System	$recall$	$precision$	EER	CS_R	AI_R
IS	80	3	18	–	–
$BS_{optDistOH}$	75	22	47	–	–
BS_{SA}	81	35	53	51	0

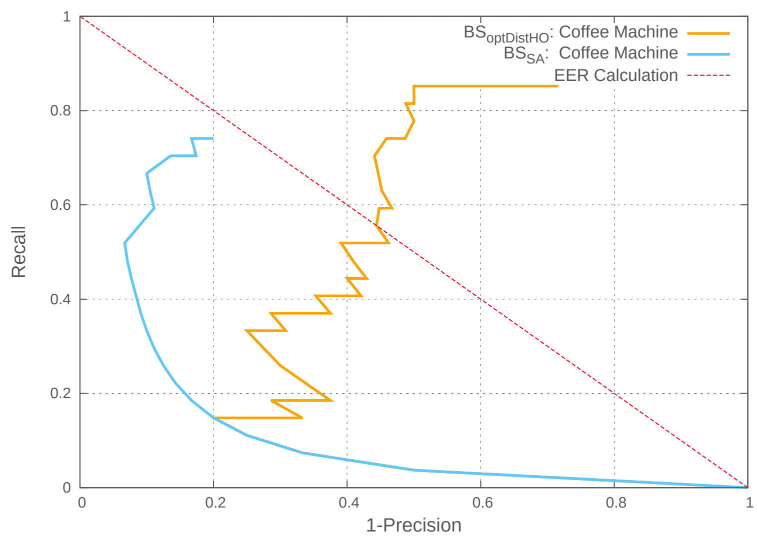
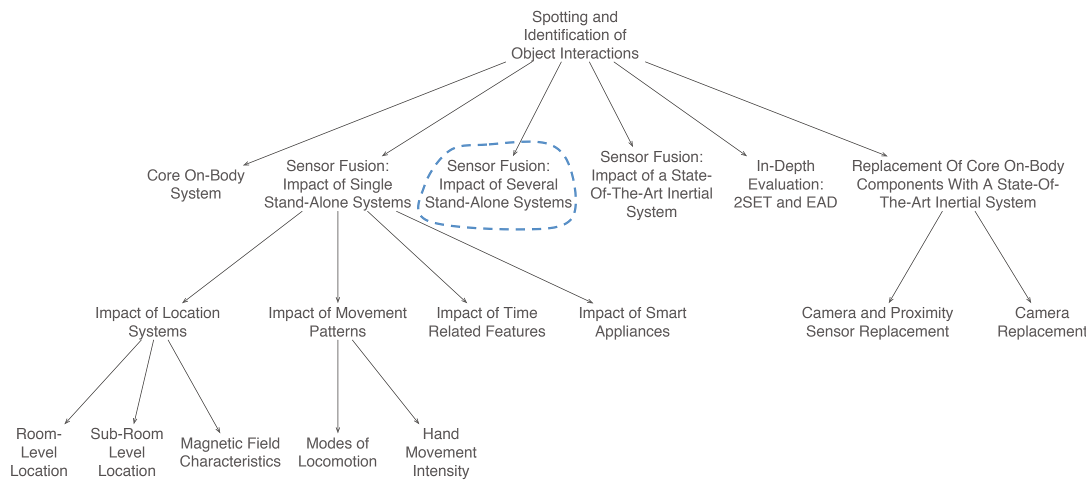


Figure 4.45: $(1 - precision_{obj}) - recall_{obj}$ curve for Coffee Machine: $BS_{optDistHO}$ (orange curve) and BS_{SA} (light blue curve).

4.7.6 Sensor Fusion – Impact of Several Stand-Alone Systems



So far the impact of stand-alone sensor systems on the recognition quality of the optimized basic system $BS_{optDistOH}$ was analyzed. Table 4.25 summarizes its results again. The best result is achieved when combining sub-room level information with $BS_{optDistOH}$. This way the *recall* can be improved by 15% to 90%, the *precision* by 18% to 40% and the EER by 14% to 61%. Compared to IS_{opt}, the *recall* can be improved by 10% and *precision* by a significant 37%. BS_{TF}, BS_{HM} and BS_{MoL} were not able to increase the recognition quality at all. However, these systems are able to reduce the amount of classification steps and analyzed images. Consequently, they contribute to a better system performance as they lower the processing time.

Table 4.25: System comparison: *recall*, *precision*, EER and percentage reduction of classification steps (CS_R) as well as analyzed images (AI_R) compared to $BS_{optDistOH}$ (optimized basic system).

System	*recall*	*precision*	EER	CS_R	AI_R
IS	80	3	18	–	–
$BS_{optDistOH}$	75	22	47	–	–
BS_{BT}	82	31	56	61	4
BS_{ROI}	90	40	61	80	28
BS_{RFL}	78	29	55	66	19
BS_{MoL}	75	22	47	3	3
BS_{HM}	75	22	46	35	42
BS_{TF}	75	22	47	14	21
BS_{SA}	81	35	53	51	0

This section evaluates the effect of fusing the optimized basic system with *multiple* stand-alone systems that were introduced so far. The combination is done by applying systems to $BS_{optDistOH}$ consecutively.

4.7.6.1 Sensor Fusion: Location Systems

First, different combinations of the location systems introduced in Section 4.7.2 are analyzed. Figure 4.46 shows the related $(1 - precision) - recall$ curves. Every system combination is able to provide a higher EER than BS_{ROI}, which was the best stand-alone system so far. However, none of the system combinations can uphold the $recall$ of 90%. The best $recall$ value decreased by 3% and was achieved by BS_{BT+ROI} and $BR_{ROI+RFL}$. There, $BR_{ROI+RFL}$ is able to provide a $precision$ of 56% which is 8% higher than was achieved by BS_{BT+ROI} and 16% higher than was reached by BS_{ROI}. However, the EER decreases by 1% to 68% compared to $BS_{BT+ROI+RFL}$, but is still 7% higher than the EER of BS_{ROI}. The highest $precision$ of 60% and the highest EER of 69% is delivered by the combination of all the location systems introduced ($BS_{BT+ROI+RFL}$). Unfortunately the $recall$ decreases by 3% compared to $BR_{ROI+RFL}$ and by 6% compared to BS_{ROI}.

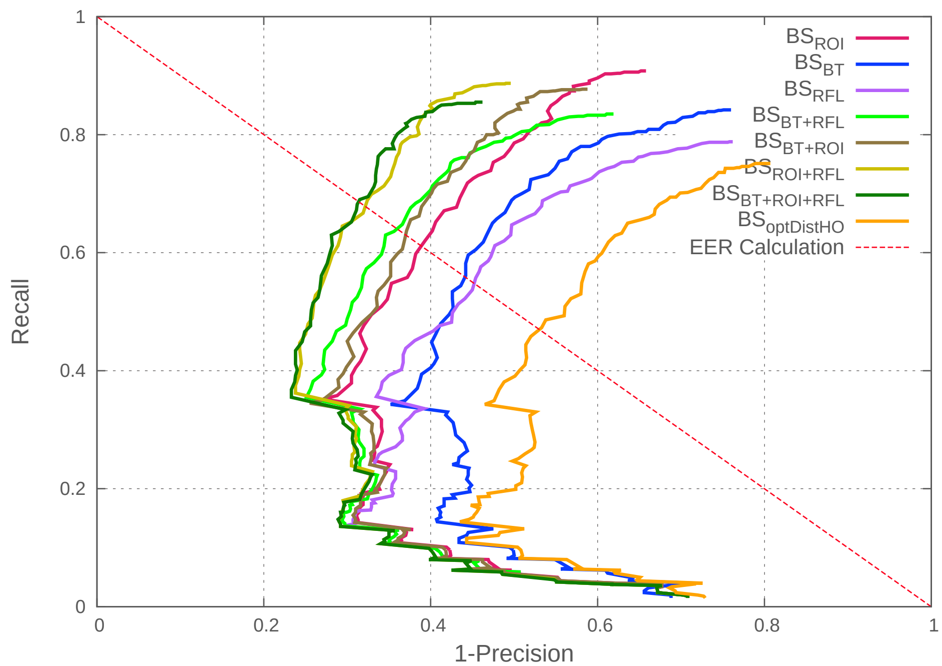


Figure 4.46: $(1 - precision) - recall$ curves (based on $thr_{svmScore}$): BS_{ROI} (red curve; optimized basic system+sub-room level location), BS_{BT} (blue curve; optimized basic system+room level location), BS_{RFL} (purple curve; optimized basic system+forearm location), BS_{BT+RFL} (light green curve; optimized basic system+room level location+forearm location), BS_{BT+ROI} (brown curve; optimized basic system+room level location+sub-room level location), $BT_{ROI+RFL}$ (yellow curve; optimized basic system+sub-room level location+forearm location), $BS_{BT+ROI+RFL}$ (dark green curve; optimized basic system+room level location+sub-room level location+forearm location) and $BS_{optDistHO}$ (orange curve; optimized basic system).

Table 4.26 shows the EER_{obj} improvement for each system combination and object compared to $BS_{optDistOH}$. Almost all combinations were able to significantly increase the EER_{obj} for each object. Only $\triangle EER_{IV}$ delivers an immense EER_{obj} decrease of 70% for "Ring Binder". As Figure 4.47 shows, from a mathematical point of view the EER_{obj} went down to zero, but in fact $precision_{obj}$ as well as $recall_{obj}$ were significantly improved even for this object. Besides, no EER_{obj} improvement was achieved for "Water Tap" and only a slight improvement was reached for "Cupboard" and "Coffee Machine". Table 4.27 summarizes the results of the fused location systems again.

Table 4.26: Object overview: Achieved EER_{obj} improvements compared to $BS_{optDistOH}$. I = BS_{BT+RFL}; II = BS_{BT+ROI}; III = $BS_{ROI+RFL}$; IV = $BS_{BT+ROI+RFL}$

Object	$\triangle EER_I$	$\triangle EER_{II}$	$\triangle EER_{III}$	$\triangle EER_{IV}$
Battery Charger	7	10	14	10
Coffee Machine	5	0	5	5
PC	6	0	24	22
Air Conditioner	78	82	91	87
Climatic Control Panel	29	33	52	29
Microwave	13	19	18	19
Ethernet Connector	27	8	34	37
Ring Binder	25	25	25	-70
Power Socket	0	12	15	15
Laser Printer	55	55	64	66
Ink Printer	22	37	35	39
Light-Shutter Switch	48	52	70	70
Scanner	20	7	22	22
Wall Cupboard	38	38	36	38
Cupboard	2	2	2	2
Water Tap	0	0	0	0
ØAverage	23	24	32	24

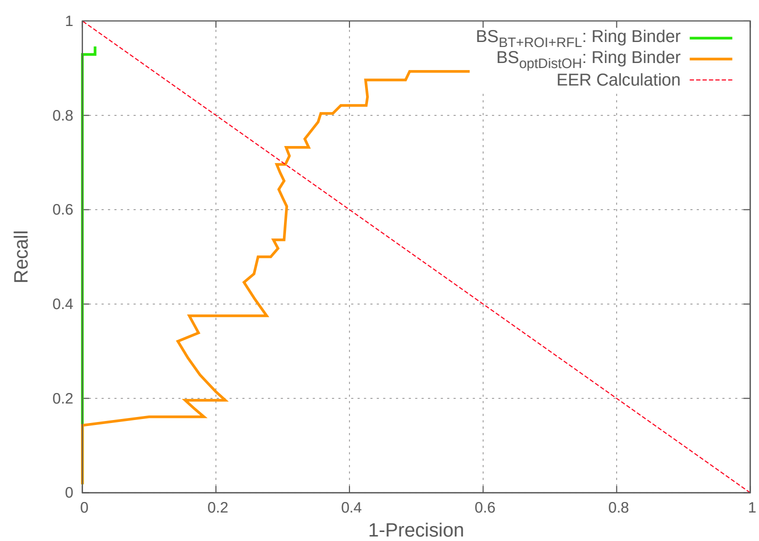


Figure 4.47: $(1 - precision_{obj}) - recall_{obj}$ curves (based on $thr_{svmScore}$) for Ring Binder: $BS_{optDistHO}$ (orange curve; optimized basic system) and $BS_{BT+ROI+RFL}$ (light green curve; optimized basic system+room level location+sub-room level location+forearm location).

As the idea was to get the highest possible *recall* and to improve the corresponding *precision* by search space restrictions with the help of additional sensor fusion approaches, only $BR_{ROI+RFL}$ was considered for further evaluations. Besides, $BR_{ROI+RFL}$ also reached the highest average object EER (see Table 4.26).

Table 4.27: System comparison (Fusion of location systems): $recall$, $precision$, EER and percentage reduction of classification steps (CS_R) as well as analyzed images (AI_R) compared to $BS_{optDistOH}$ (optimized basic system).

System	$recall$	$precision$	EER	CS_R	AI_R
IS	80	3	18	–	–
$BS_{optDistOH}$	75	22	47	–	–
BS_{BT}	82	31	56	61	4
BS_{ROI}	90	40	61	80	28
BS_{RFL}	78	29	55	66	19
BS_{MoL}	75	22	47	3	3
BS_{HM}	75	22	46	35	42
BS_{TF}	75	22	47	14	21
BS_{SA}	81	35	53	51	0
BS_{BT+ROI}	87	48	63	84	38
BS_{BT+RFL}	82	46	64	85	36
$BS_{ROI+RFL}$	87	56	68	90	49
$BS_{BT+ROI+RFL}$	84	60	69	91	53

4.7.6.2 Sensor Fusion: Movement Patterns, Time Related Features and Smart Appliances

As was already shown, none of the introduced features related to movement patterns or time were able to increase *recall*, *precision* or EER. Consequently, sensor fusion approaches based on these features will not influence the outcome of the basic system either. Nevertheless, the amount of classification steps and analyzed images could be reduced. Due to this fact, features weren't combined with each other but with the location systems and smart appliances. As BS_{HM} (hand movement intensity) is even worse than $BS_{optDistOH}$ with respect to EER, this feature is excluded from further evaluations. Table 4.28 shows the analyzed fusion approaches and their corresponding results.

Table 4.28: System comparison (Overall system fusion): *recall*, *precision*, EER and percentage reduction of classification steps (CS_R) as well as analyzed images (AI_R) compared to $BS_{optDistOH}$ (optimized basic system).

System	*recall*	*precision*	EER	CS_R	AI_R
IS	80	3	18	–	–
$BS_{optDistOH}$	75	22	47	–	–
BS_{BT}	82	31	56	61	4
BS_{ROI}	90	40	61	80	28
BS_{RFL}	78	29	55	66	19
BS_{MoL}	75	22	47	3	3
BS_{HM}	75	22	46	35	42
BS_{TF}	75	22	47	14	21
BS_{SA}	81	35	53	51	0
BS_{BT+ROI}	87	48	63	84	38
BS_{BT+RFL}	82	46	64	85	36
$BS_{ROI+RFL}$	87	56	68	90	49
$BS_{BT+ROI+RFL}$	84	60	69	91	53
$BS_{TF+MoL+RFL}$	78	30	55	70	32
$BS_{TF+MoL+ROI+RFL}$	86	57	68	91	56
$BS_{SA+ROI+RFL}$	90	70	79	94	63
$BS_{TF+MoL+SA+ROI+RFL}$	89	70	79	95	68

The highest *recall* was achieved by combining the basic system with smart appliances (SA), a sub-room level location system (ROI) and the user's hand location (RFL). This way, a *recall* of 90%, a corresponding *precision* of 70% and an EER of 79% was accomplished. Besides, the amount of classification steps was significantly reduced by 94% – so has the number of analyzed images (by 63%). When adding time related features (TF) and modes of locomotion (MoL), similar results can be achieved. As it was expected, there is a higher reduction of classification steps as well as analyzed images. Nevertheless, the *recall* decreases by 1% to 89%. If smart appliances are not available, the highest *recall* is still achieved by $BT_{ROI+RFL}$. As before, the combination of time related features and modes of locomotion reduces the amount of analyzed images, but involves a slight *recall* decrease (1%). However, this results in a slight *precision* increase of 1%.

In some applications it is neither possible to use smart appliances nor to install any other sensor systems (e.g. cameras). In such scenarios $BT_{TF+MoL+RFL}$ can be used as it is exclusively based on wearable sensors attached to the user's body. Consequently, there is no need for any environmental instrumentation. Compared to $BT_{SA+ROI+RFL}$, which is so far the best ranked combination, the *recall* was decreased by 12% to 78%, the *precision* by 40% to 30% and the EER by 24% to 55%. Figure 4.48 shows the respective $(1-precision)-recall$ curves. Looking at specific objects and EER_{obj} this conclusion can be confirmed (see Table 4.29). The best result was achieved when using $BS_{SA+ROI+RFL}$. The average EER_{obj} was increased by 35% compared to $BS_{optDistOH}$. The worst improvement of 10% was reached when using body-worn

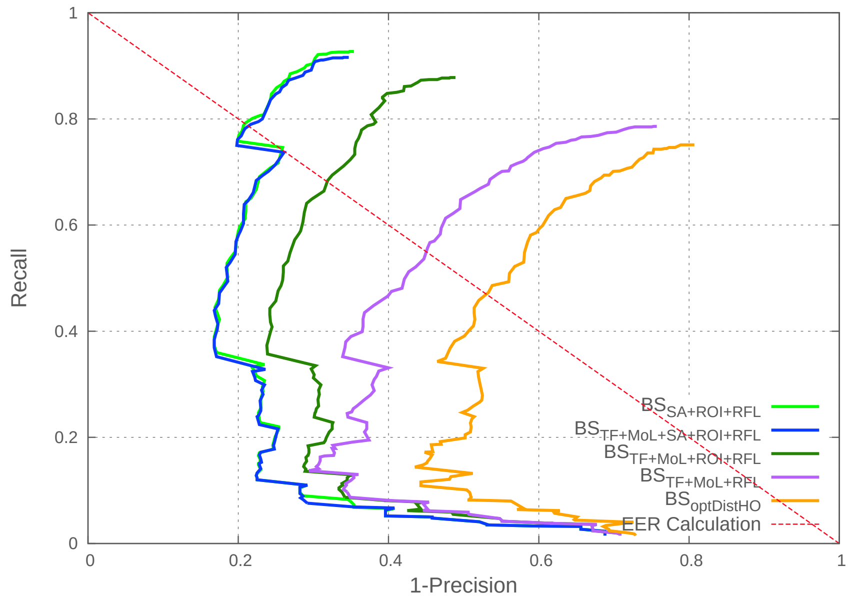


Figure 4.48: $(1-precision)-recall$ curves (based on $thr_{svmScore}$): $BS_{SA+ROI+RFL}$ (light green curve; optimized basic system+smart appliances+sub-room level location+forearm location), $BS_{TF+MoL+SA+ROI+RFL}$ (blue curve; optimized basic system+temporal feature+mode of locomotion+smart appliances+sub-room level location+forearm location), $BS_{TF+MoL+ROI+RFL}$ (dark green curve; optimized basic system+temporal feature+mode of locomotion+sub-room level location+forearm location), $BS_{TF+MoL+RFL}$ (purple curve; optimized basic system+temporal feature+mode of locomotion+forearm location) and $BS_{optDistOH}$ (orange curve; optimized basic system).

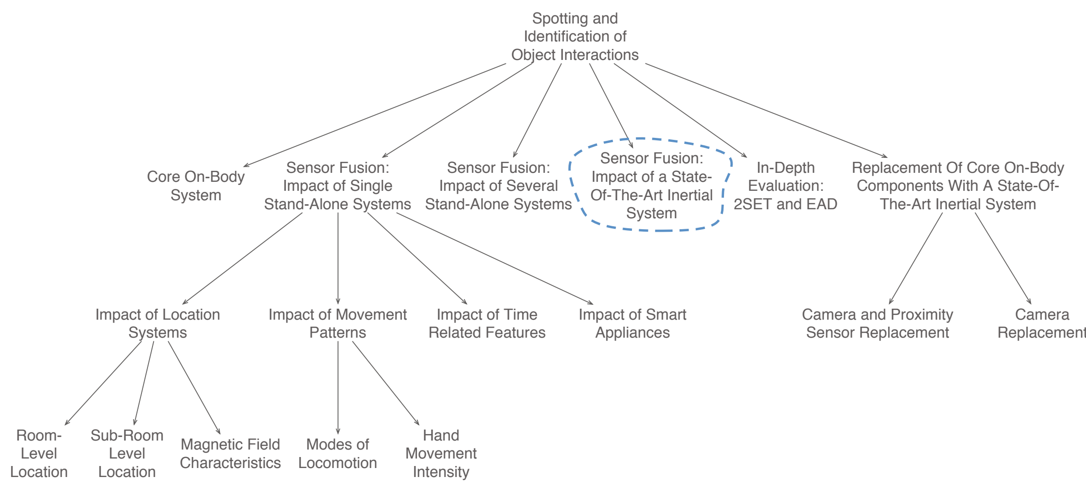
sensors only. However, the infrastructure was not instrumented in this scenario. In the following, further evaluations are based on three system configurations:

- $BS_{SA+ROI+RFL}$ (optimized basic system + smart appliances + sub-room level location + forearm location). This approach achieved the best results.
- $BS_{ROI+RFL}$ (optimized basic system + sub-room level location + forearm location). This is the best system configuration without smart appliances.
- $BS_{TF+MoL+RFL}$ (optimized basic system + temporal features + modes of locomotion + forearm location). This approach reached the best results without environmental instrumentations.

Table 4.29: Object overview: Achieved EER_{obj} improvements of $\triangle EER_I - \triangle EER_{IV}$ (I = $BS_{SA+ROI+RFL}$; II = $BS_{TF+MoL+SA+ROI+RFL}$; III = $BS_{TF+MoL+RFL}$; IV = $BS_{TF+MoL+ROI+RFL}$) compared to $BS_{optDistOH}$.

Object	$\triangle EER_I$	$\triangle EER_{II}$	$\triangle EER_{III}$	$\triangle EER_{IV}$
Battery Charger	10	10	0	10
Coffee Machine	29	29	8	7
PC	87	0	0	26
Air Conditioner	0	0	62	89
Climatic Control Panel	48	48	9	52
Microwave	40	40	8	18
Ethernet Connector	30	30	7	34
Ring Binder	25	25	17	-70
Power Socket	15	15	-44	15
Laser Printer	64	66	3	66
Ink Printer	32	37	3	39
Light-Shutter Switch	70	70	44	70
Scanner	34	34	5	23
Wall Cupboard	36	36	33	36
Cupboard	2	2	2	2
Water Tap	44	44	0	0
ØAverage	35	30	10	26

4.7.7 Sensor Fusion – Impact of a State-Of-The-Art Inertial System



State-of-the-art approaches use inertial sensors and large training data sets in order to spot and to recognize human activities. Section 4.3 has already shown, that such a system does not provide sufficient recognition rates for the problem considered. However, in this section the introduced system was combined with a trained inertial sensor system. Of course, this fact prohibits an easy large-scale deployment in real-life scenarios, which is one of the key requirements in the scope of this thesis. Nevertheless, it should be investigated whether information coming from inertial sensors and large training data sets can significantly improve results achieved so far.

4.7.7.1 Integration Procedure

The following steps describe the fusion procedure:

- Spot interesting segments TS_{IS_i} for each activity using a trained inertial sensor system (see Section 4.3.1).
- Calculate a ranked list ($List_{IS_{rankedObj}}$) containing potential object interactions based on recognized activities for each TS_{IS_i}.
- Check for all ranked object candidates in TSS_{i_j} (created by BS) in a consecutive order (starting with the highest ranked object) if it is also included in $List_{IS_{rankedObj}}$. The first object which is included in both lists is used as final winner. If such an object is not present, it is assumed that no activity or object interaction took place within TSS_{i_j}.

Figure 4.49 visualizes the fusion process again. The idea is to use a configuration of IS which is able to reach the highest *recall* in order to minimize false classifications detected by BS. As was shown in Section 4.3.3, the highest *recall* (80%) was achieved for $IS_{thrSVMScore} = 0.0135$ and $IS_{thrFusion} = 0$ seconds. Hence, this configuration was considered again. Table 4.30 and Table 4.31 show results for the fused systems $BS_{SA+ROI+RFL} + IS$, $BS_{ROI+RFL} + IS$ and $BS_{TF+MoL+RFL} + IS$. It can be seen, that no improvements can be reached with respect to *precision*, EER and the amount of analyzed images as well as classification steps. On the

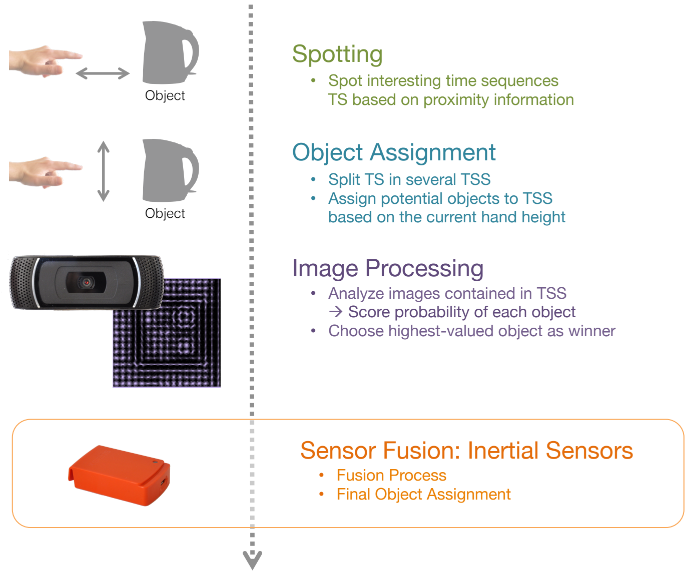


Figure 4.49: BS: Classification Procedure.

contrary, a slight *recall* decrease results from the fusion with IS. Looking at specific objects, this fact can be partially confirmed. In terms of $BS_{SA+ROI+RFL}+IS$ and $BS_{ROI+RFL}+IS$ the average EER_{obj} decreases by 6% and 12%. It is worth noting, that in both cases the EER_{obj} for "Ring Binder" decreases by an immense 95%. Besides, $BS_{ROI+RFL}+IS$ shows an EER_{obj} decrease for "Air Conditioner" of 91%. As Figure 4.50 shows, with respect to these configurations the EER_{obj} went down to zero from a mathematical point of view. But as $(1-precision_{obj})-recall_{obj}$ curves show, combining BS with inertial sensors cannot improve the results for these specific objects at all. Only $BS_{TF+MoL+RFL}+IS$ is able to increase the average EER_{obj} by 4%. There, significant improvements could be achieved in case of "Ethernet Connector", "Power Socket", "Light-Shutter Switch" and "Water Tap". Other objects could uphold their EER_{obj} or showed only a slight decrease.

Table 4.30: System comparison (System combinations with IS): Achieved improvements compared to particular systems without their fusion with IS and percentage reduction of classification steps (CS_R) as well as analyzed images (AI_R) compared to $BS_{optDistOH}$ (optimized basic system)

System	$\triangle Recall$	$\triangle Precision$	$\triangle EER$	$\triangle CS$	$\triangle AI$
$BS_{SA+ROI+RF}+IS$	-1	0	0	0	0
$BS_{ROI+RFL}+IS$	-2	0	0	0	0
$BS_{TF+MoL+RFL}+IS$	-3	0	0	0	0

In summary we can see, that the combination of both systems is not able to improve the overall system recognition quality. Only a small improvement can be achieved in case of $BS_{TF+MoL+RFL}+IS$. This fact again confirms, that a state-of-the-art approach based on inertial sensors is neither able to solve the recognition problem in focus nor improve recognition qualities reached so far.

Table 4.31: Object overview: Achieved EER_{obj} improvements compared to particular systems $\triangle EER_I - \triangle EER_{III}$ (I $= BS_{ROI+RFL} + IS$; II $= BS_{SA+ROI+RFL} + IS$ and III $= BS_{TF+MoL+RFL} + IS$) without their fusion with IS.

Object	$\triangle EER_I$	$\triangle EER_{II}$	$\triangle EER_{III}$
Battery Charger	-4	0	-4
Coffee Machine	0	0	3
PC	0	0	0
Air Conditioner	-91	0	-5
Climatic Control Panel	0	0	0
Microwave	0	0	-1
Ethernet Connector	0	0	10
Ring Binder	-95	-95	-1
Power Socket	0	0	44
Laser Printer	-2	-2	4
Ink Printer	0	0	-4
Light-Shutter Switch	2	2	10
Scanner	-1	-1	1
Wall Cupboard	2	2	2
Cupboard	-4	-4	-4
Water Tap	0	0	11
ØAverage	-12	-6	4

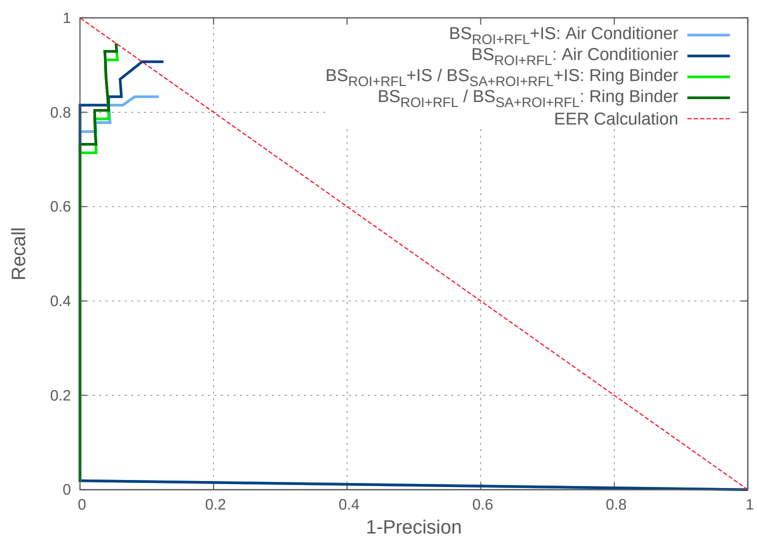
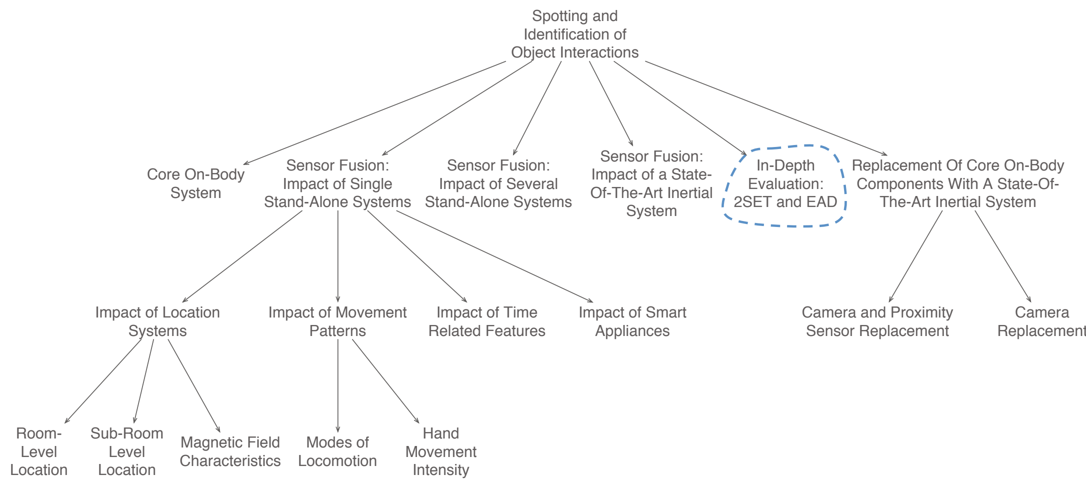


Figure 4.50: $(1 - precision_{obj}) - recall_{obj}$ curves (based on $thr_{svmScore}$): "Air Conditioner": $BS_{ROI+RFL} + IS$ (light blue curve; optimized basic system + sub-room level location+forearm location + inertial system), $BS_{ROI+RFL}$ (blue curve; optimized basic system + sub-room level location + forearm location). "Ring Binder": $BS_{ROI+RFL} + IS$ (light green curve; optimized basic system+sub-room level location+forearm location+inertial system), $BS_{SA+ROI+RFL} + IS$ (light green curve; optimized basic system + smart appliances + sub-room level location + forearm location + inertial system), $BS_{ROI+RFL}$ (green curve; optimized basic system + sub-room level location + forearm location) and $BS_{SA+ROI+RFL}$ (green curve; optimized basic system + smart appliances + sub-room level location + forearm location). "Ring Binder" shows exactly the same $(1 - precision_{obj}) - recall_{obj}$ curve in terms of evaluated systems when using and not using smart appliances.

4.7.8 In-Depth Evaluation: 2SET and EAD



So far, a standard evaluation approach based on precision and recall was used in order to benchmark the systems introduced. However, using only false positives, true positives and false negatives on an event level might not be sufficient to evaluate systems dealing with continuous activity recognition problems as is considered in this work. Common artifacts such as event fragmentations or merges, which normally occur in spotting problems, are not covered. To overcome this problem, a more detailed evaluation procedure, that is introduced in [WLG11], was applied, whereby two evaluation methods based on frames and events are described. The following evaluations are based on three system fusion approaches that achieved the best recognition accuracy for the kind of spotting problem considered:

- The system achieving the best overall recognition rates: $BS_{SA+ROI+RFL}$ (The core system in combination with right forearm and sub-room level location as well as smart appliances; EER of 79%)
- The system achieving the best recognition rates in applications where minimal infrastructure effort is of interest: $BS_{ROI+RFL}$ (The core system in combination with right forearm and sub-room level location; EER of 68%)
- The system achieving the best recognition rates in applications where infrastructure instrumentation is not possible: $BS_{TF+MoL+RFL}$ (The core system in combination with right forearm location, modes of locomotion and time features; EER of 55%)

4.7.8.1 Frame Based Evaluation – 2 SET

The first method evaluates the system on a frame level. A frame is defined as a fixed-length, fixed-rate unit of time. In many cases it is the smallest unit of measure which is defined by the system – hence it approximates continuous time. In this work, a frame is defined as one second. As described in [WLG11] the frame-based evaluation is represented by a so-called two-class segment error table (2SET) which considers deletions (dr), fragmentations (fr), start underfill (u^{α}), end underfill (u^{ω}) and true positive rate (tpr) as well as insertions (ir), merges (mr), start overfill (o^{α}), end overfill (o^{ω}) and false positive rate (fpr). These evaluation features where calculated for each object as described in [WLG11]. In order to get an overall system evaluation, a 2SET representation based on average evaluation features of each object was used. Figure 4.51 to Figure 4.53 visualize 2SETs of $BS_{SA+ROI+RFL}$, $BS_{ROI+RFL}$ and $BS_{TF+MoL+RFL}$. It can be seen, that for each system the amount of u^{α}, u^{ω} and fr is insignificantly small compared to real deletions. Consequently, tpr confirms the recall values already achieved. On the contrary, quite different results are achieved in terms of false positive rates. Each system delivers a rather low false positive rate (below 3%). This means, that almost all background frames (real background frames and frames belonging to objects other than the one considered) were correctly recognized. However, as the precision values were not outstanding (between 30% and 70%), it can be seen, that the amount of o^{α}, o^{ω}, mr and ir is insignificantly small compared to the amount of background frames. This means, that even if the number of insertions is relatively high compared to the amount of activities performed, they are very low in respect of time frames. Consequently, all systems investigated are able to sort out almost all background frames. This fact was not discovered by previously used evaluation techniques.

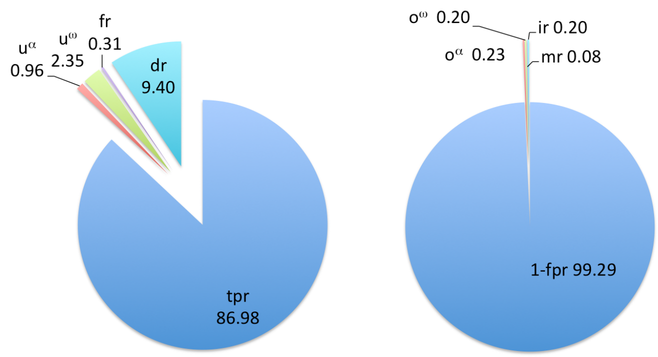


Figure 4.51: 2SET representation: $BS_{SA+ROI+RFL}$.

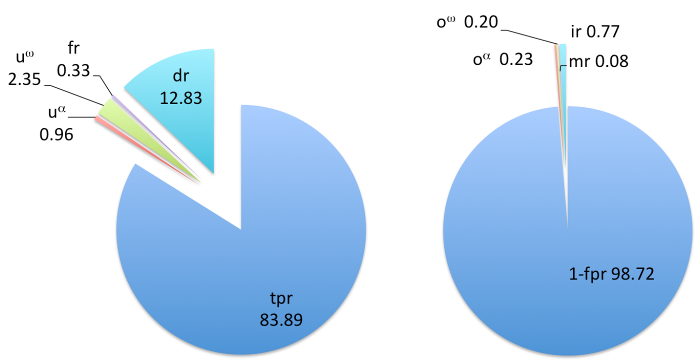


Figure 4.52: 2SET representation: $BS_{ROI+RFL}$.

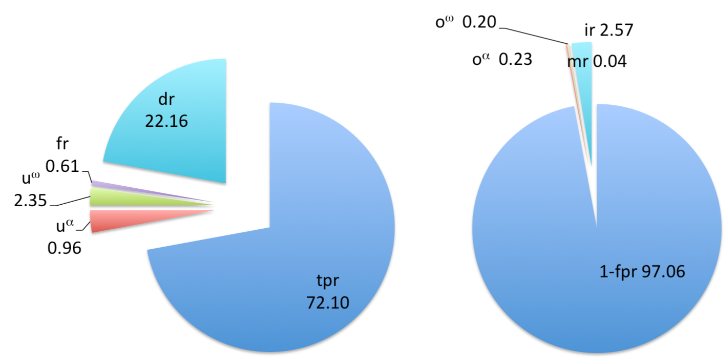


Figure 4.53: 2SET representation: $BS_{TF+MoL+RFL}$.

4.7.8.2 Event Based Evaluation – EAD

Apart from a frame level evaluation, Ward proposes a benchmarking technique based on events in [WLG11]. Events are defined as time sequences which consist of positive frames. They have a variable duration and are defined by a start and end time. Two event types are considered: The first one is a set of events $E = \{e_1, e_2, ..., e_g\}$ describing activities, that were performed by the user and therefore contain the so called "ground truth". The other event set $R = \{r_1, r_2, ..., r_h\}$ includes the recognition output of a system. As there is not necessarily a one-to-one relationship between E and R, events can be scored as: correctly recognized (C), falsely inserted (I'; there is no corresponding event in the ground truth) or deleted (D; a performed event $e_i \epsilon E$ was not detected). Based on these definitions, various benchmarking features such as recall, precision, true positive rate or false positive rate can be calculated. So far [WLG11] describes an event-based benchmarking technique which complies with our evaluation procedure, that was previously introduced. In order to get a more precise evaluation, [WLG11] extends the list of event categories (events included in E / R) by fragmentation (F / F'), merge (M / M') as well as fragmentation and merge (FM / FM'). All features are visualized in a single event analysis diagram (EAD), which shows the number of all event categories. To make it comparable across variably sized datasets, all numbers are reduced to percentages: $|E|$, $\frac{C}{|E|}$, $\frac{D}{|E|}$, $\frac{F}{|E|}$, $\frac{M}{|E|}$ and $\frac{FM}{|E|}$ as well as $|R|$, $\frac{C}{|R'|}$, $\frac{I'}{|R'|}$, $\frac{F'}{|R|}$, $\frac{M'}{|R|}$ and $\frac{FM'}{|R|}$. Figure 4.54 visualizes EAD of $BS_{SA+ROI+RFL}$, $BS_{ROI+RFL}$ and $BS_{TF+MoL+RFL}$. It is noticeable, that the evaluation is not based on an average object level as was done before. Instead all features were calculated based on the sum of each object-related feature (e.g. $I' = \sum_{m=1}^{n} I'_m$ with $n = \#objects$ and $I'_m = Insertions_{Object_m}$).

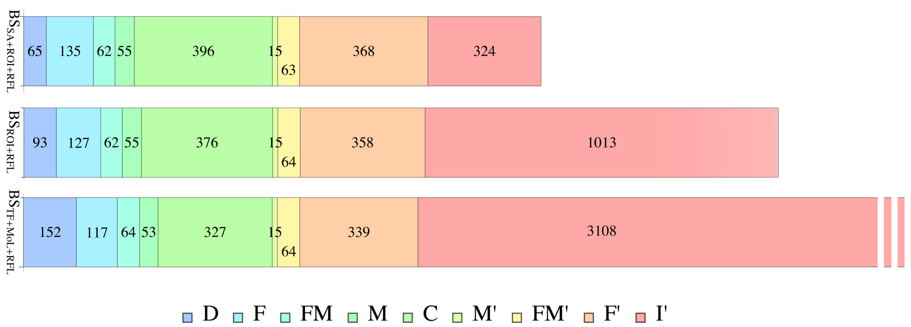


Figure 4.54: EAD: $BS_{SA+ROI+RFL}$, $BS_{ROI+RFL}$ and $BS_{TF+MoL+RFL}$.

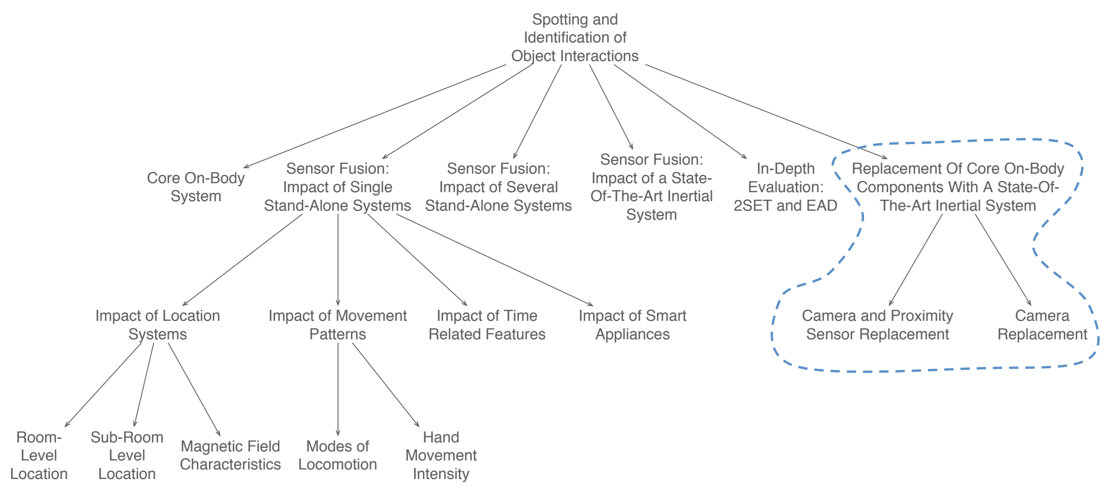
As could be expected due to the previous evaluations, there is a strong increase related to the quantity of deleted events D starting from $BS_{SA+ROI+RFL}$ to $BS_{TF+MoL+RFL}$ (see Figure 4.54). The number of F, M and FM is nearly the same for all considered systems and is between 32% and 36%. There, the amount of F is almost equal to the sum of M and FM. Looking at events classified as true, it can be seen that between 46% and 56% of the events were correctly recognized. This means, that for each system more than half of all objects classified as true were in fact correctly recognized and not merged, fragmented or both (see Table 4.32).

Table 4.32: EAD (C, F, FM, M) for $BS_{SA+ROI+RFL}$, $BS_{ROI+RFL}$ and $BS_{TF+MoL+RFL}$ in percent (rounded values) based on performed events E.

System	C	F	FM	M	D
$BS_{SA+ROI+RFL}$	56	8	9	19	9
$BS_{ROI+RFL}$	53	8	9	18	13
$BS_{TF+MoL+RFL}$	46	7	9	16	21

Looking at F', FM' and M', all systems have a rather static behavior. However, the number of insertions shows enormous variations (see Figure 4.54). There, $BS_{SA+ROI+RFL}$ has about three times less insertions than $BS_{ROI+RFL}$ and nearly ten times less insertions than $BS_{TF+MoL+RFL}$. This fact illustrates the significant improvements, that were achieved by sensor fusion techniques, again. As this evaluation approach is not based on an average object level, further comparisons between the achieved precision/recall values and previously reached results were not considered.

4.7.9 Replacement of Core On-Body System Components with a State-Of-The-Art Inertial System



The objective of this section is to evaluate the impact of replacing components of the core on-body systems with a motion system. This idea is motivated by the following facts:

- A minimal core on-body sensor setup should be focused on. Therefore, the camera and the proximity sensor were removed from the core system. Consequently, the remaining human body model and hand height information derived are fused with an inertial system using a statistically significant, large training set.
- The motion system was able to achieve a rather high *recall* of 80% but a very poor corresponding *precision* of only 3% and a EER of 18% for the considered problem. This means, that IS delivers a large amount of falsely classified activities and resulting object interactions. The on-body system applies a vision-based object recognition method on top of a quite restricted search space. Hence, the impact of replacing the wrist camera and vision-based processing methods of the core system with a motion system should be evaluated.

Both concepts are described more specifically in the following.

4.7.9.1 Camera and Proximity Sensor Replacement

First, the camera and the proximity sensor were removed from the core system. Hence, information about the person's hand height derived from the human body model of the core system was combined with IS. In the following, this system is referred to as IS'_{HH}. The "Final Segmentation Procedure" step (see Section 4.3.1) was replaced (and consequently the spotting process of IS) with the following steps based on ideas of the core on-body system:

- Interesting sequences $IS'_{HH_{TS_i}}$ were defined based on the current hand height variation. For each $IS'_{HH_{TS_i}}$ the maximum hand height variation allowed is less than 10 cm. Therefore exactly the same procedure as was used for defining TSS_{i_j} in $BS_{optDistOH}$ was applied.

- For each $IS'_{HH_{TS_i}}$ the average hand height was calculated. A list containing potential objects was created based on the average hand height, the height of each object and a deviation threshold of $(thrDistHO_{up}, thrDistHO_{down}) = (25\text{ cm}, 10\text{ cm})$ (which was already optimized for BS) .
- For each object contained the average SVM score (calculated based on timeframes as described in Section 4.3.1) of $IS'_{HH_{TS_i}}$ was determined.
- The object interaction with the highest SVM score was chosen as performed object interaction in $IS'_{HH_{TS_i}}$.
- Finally, a SVM score threshold ($IS'_{HH_{SVMThr}}$) was used to reject low-ranked results.

The impact of $IS'_{HH_{SVMThr}}$ (ranging from 0 to 1 in steps of 0.001) on *precision* and *recall* was evaluated. The highest *recall* of 32% and a corresponding *precision* of 3% was achieved with $IS'_{HH_{SVMThr}} = 0$ (This means, that in fact no object findings were rejected). Consequently, the new fusion approach reduced the *recall* by 43% and the corresponding *precision* by 19%. Figure 4.55 visualizes $(1 - precision) - recall$ curves and the reached EER. It can be seen, that when using information about the person's hand height, the adopted final segmentation procedure delivers even worse results than IS and therefore much worse results than $BS_{optDistOH}$. This is also confirmed by Table 4.33, which shows the recognition accuracy for specific objects. In fact, a strong decrease can be seen for both the average $recall_{obj}$ and the average $precision_{obj}$. It is worth noting, that the $recall_{obj}$ for the "Scanner" device can be significantly increased by 10%, whereas almost all other devices show an immense $recall_{obj}$ reduction. In summary, we can see that the proposed method based on average SVM scores cannot reach the results achieved so far.

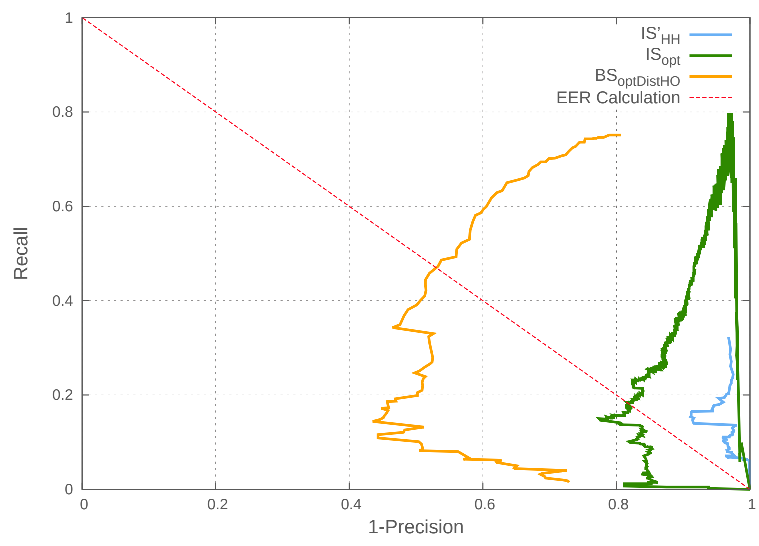


Figure 4.55: $(1 - precision) - recall$ curves (based on $thr_{svmScore}$): IS'_{HH} (light blue curve; inertial system+hand height information), IS_{opt} (green curve; inertial system) and $BS_{optDistOH}$ (orange curve; optimized basic system).

Table 4.33: Object overview for IS'_{HH}: $recall_{obj}$, $precision_{obj}$, EER_{obj} and improvements compared to $BS_{optDistOH}$ in terms of $recall_{obj}$ ($\triangle Rec$), $precision_{obj}$ ($\triangle Prec$) and EER_{obj} ($\triangle EER$).

Object	$recall_{obj}$	$precision_{obj}$	EER_{obj}	$\triangle Rec$	$\triangle Prec$	$\triangle EER$
Battery Charger	0	0	0	-83	-54	-76
Coffee Machine	7	2	7	-78	-29	-49
PC	0	0	0	0	0	0
Air Conditioner	0	0	0	-43	-60	0
Climatic Control Panel	33	1	4	-49	-25	-37
Microwave	38	3	4	-61	-5	-44
Ethernet Connector	70	2	3	-3	-8	-50
Ring Binder	27	5	5	-62	-44	-65
Power Socket	59	1	4	-11	-27	-40
Laser Printer	7	0	0	-56	-2	-19
Ink Printer	11	1	7	-82	-1	-4
Light-Shutter Switch	39	1	2	-29	-6	-12
Scanner	98	2	2	3	-5	-23
Wall Cupboard	91	6	15	10	-8	-11
Cupboard	33	26	21	-45	-17	-55
Water Tap	4	1	4	-92	-4	-26
ØAverage	32	3	5	-42	-19	-34

4.7.9.2 Camera Replacement

As a next step, exactly the same approach was used to spot interesting sequences and to restrict the number of potential object interactions as was already used for BS. This means, a proximity sensor was used to spot interesting sequences and potential objects were chosen based on the average hand height within such a sequence. Only the final object recognition step was replaced by an inertial approach. In the following this system is referenced with $IS'_{PROX+HH}$. The following processing steps were performed:

- Spot interesting time sequences $IS'_{PROX+HH_{TS_i}}$ where the distance between the user's finger tip and an object is less than 10 cm.
- Divide each $IS'_{PROX+HH_{TS_i}}$ in several $IS'_{PROX+HH_{TS_{i_j}}}$. The maximum hand height variation allowed within each $IS'_{PROX+HH_{TS_{i_j}}}$ is 10 cm.
- Assign a list of potential objects to each $IS'_{PROX+HH_{TS_{i_j}}}$ based on the average hand height. There, a deviation between the exact object height and the user's hand height of $(thrDistHO_{up}, thrDistHO_{down})$ = (25 cm, 10 cm) is allowed. This value was already optimized for BS.
- For each object in the list of $IS'_{PROX+HH_{TS_{i_j}}}$, the average SVM score (calculated timeframe-based as described in Section 4.3.1) is determined.
- The object interaction with the highest SVM score is chosen as performed object interaction in $IS'_{Prox+HH_{TS_i}}$.
- Finally, a SVM score threshold of $IS'_{Prox+HH_{SVMThr}}$ is used to reject low-ranked results.

The impact of various $IS'_{Prox+HH_{SVMThr}}$ (ranging from 0 to 1 in steps of 0.001) on the overall classification accuracy was analyzed. The highest *recall* of 42% with a corresponding *precision* of 9% was achieved with $IS'_{Prox+HH_{SVMThr}} = 0.004$. Again, compared to $BS_{optDistOH}$ the *recall* significantly decreased by 33% to 42%, the corresponding *precision* by 13% to 9% and the EER by 20% to 27%. Figure 4.56 shows the corresponding $(1 - precision) - recall$ curves. This fact is confirmed by analyzing the impact on the recognition quality for specific objects as shown in Table 4.34. There, the average EER_{obj} was decreased by 18%, the average $precision_{obj}$ by 13% and the $recall_{obj}$ by 32%. However, a significant EER_{obj} increase can be seen for three objects: The EER_{obj} of "Light-Shutter Switch" was increased by 16%, the EER_{obj} of "Climatic Control Panel" by 26% and the EER_{obj} of "Wall Cupboard" was increased by an immense 61%.

Evaluations have shown, that both approaches are unable to provide similar results as achieved by the core on-body system in its original form. By contrast, the replacement of core on-body components results in a significant average EER_{obj} decrease of 18% and 34%. In the following, the spotting and recognition of subtle hand activity, which were the basis of the object interactions performed, are considered.

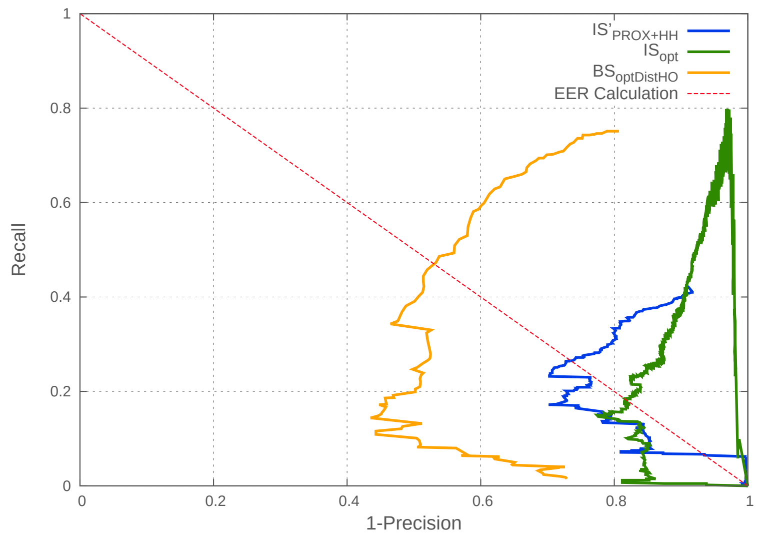


Figure 4.56: $(1 - precision) - recall$ curves (based on $thr_{svmScore}$): $IS'_{PROX+HH}$ (blue curve; inertial system+proximity information+hand height information), IS_{opt} (green curve; inertial system) and $BS_{optDistOH}$ (orange curve; optimized basic system).

Table 4.34: Object overview for $IS'_{PROX+HH}$: $recall_{obj}$, $precision_{obj}$, EER_{obj} and improvements compared to $BS_{optDistOH}$ in terms of $recall_{obj}$ ($\triangle Rec$), $precision_{obj}$ ($\triangle Prec$) and EER_{obj} ($\triangle EER$).

Object	$recall_{obj}$	$precision_{obj}$	EER_{obj}	$\triangle Rec$	$\triangle Prec$	$\triangle EER$
Battery Charger	0	0	0	-83	-54	-76
Coffee Machine	11	8	11	-74	-23	-45
PC	2	1	0	2	1	0
Air Conditioner	0	0	0	-43	-60	0
Climatic Control Panel	96	5	67	14	-21	26
Microwave	96	7	41	-3	-1	-7
Ethernet Connector	13	6	7	-60	-4	-46
Ring Binder	84	27	39	-5	-22	-31
Power Socket	41	3	11	-29	-25	-33
Laser Printer	0	0	0	-63	-2	-19
Ink Printer	29	14	14	-60	9	-33
Light-Shutter Switch	71	8	30	3	1	16
Scanner	100	4	8	5	-3	-17
Wall Cupboard	98	28	87	17	14	61
Cupboard	35	33	24	-43	-10	-52
Water Tap	0	0	0	-96	-5	-30
ØAverage	42	9	21	-32	-13	-18

4.8 Spotting and Identification of Subtle Arm Activities: System Implementation and Evaluation

So far the focus of this chapter was on spotting and identifying object interactions coming from subtle arm activities within a continuous data stream. Although this kind of information could be sufficient for many applications, other scenarios may need more detailed information related to the activities performed. Hence, this section aims at recognizing underlying subtle arm activities, being the origin of object interactions. That way more detailed information about performed object interactions should be obtained. The benefit of such systems is clear: Instead of spotting an undefined device interaction event, more detailed information about the object interaction performed such as "device opened", "device closed", "button on device pressed" or "device cleaned" is reached.

When considering related subtle hand activities, it is clear, that the systems introduced so far are unable to distinguish between specific activities such as opening, closing or turning a device on and off. The reason for this is obvious: Considered sensor modalities and corresponding data do not include helpful information needed to distinguish between such activities. Of course, using image processing algorithms, one can track signs on a devices' control panel and consequently recognize its current state (if it is displayed). However, this would require the use of complex and device-specific algorithms which would prohibit the idea of a large-scale and easy to set up system (considering the sheer volume of different devices). The idea of this section is to evaluate how detailed underlying activities can be detected by adding additional information coming from smart devices and a trained inertial sensor system to the system introduced. It is worth noting, that only object interactions with several different underlying activities were considered and single underlying activities were automatically mapped.

In the following, three fusion approaches (see Figure 4.57) considering different groups of activities are addressed. There, the introduced system $BS_{SA+ROI+RFL}$ is used in combination with additional information coming from smart devices, a trained inertial sensor system and the fusion of both systems.

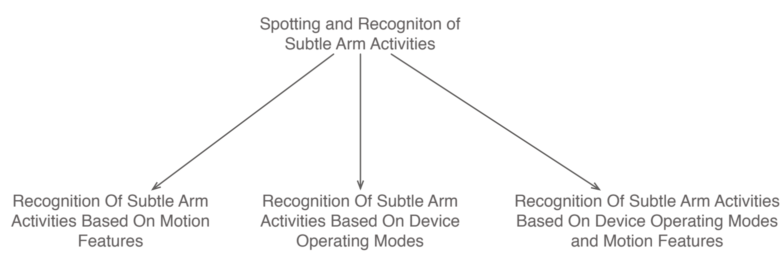
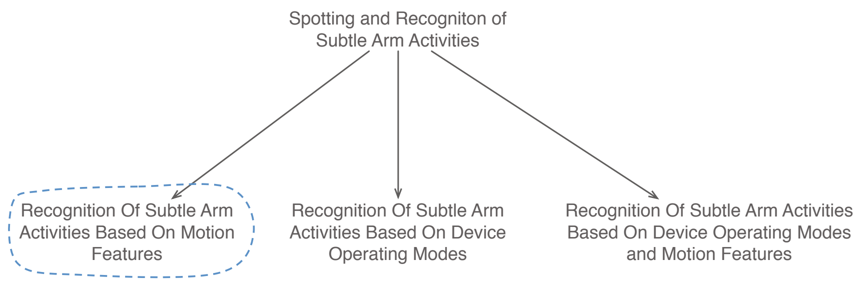


Figure 4.57: Overview: Spotting and identification of subtle arm activities.

The activity sets in focus contain between 25 and 32 activities. The reason for this is, that some object related activities were merged to a single unspecific object event as they cannot be distinguished due to the nature of the approach used (e.g. it is impossible to recognize if a button was pushed to turn the device on or off by analyzing motion patterns only). In the following, $(1-precision)-recall$ curves for the systems based on $BS_{SA+ROI+RFL}$ were created by varying $thr_{svmScore}$ from -0.4 to 1.0 in steps of 0.01 and activity related recall and precision values have been calculated for $thr_{svmScore}$ = -0.18 (which delivered the best results for BS). It is worth noting, that the trained inertial sensor systems and the corresponding sample based recognition results, which were used in this section for further fusion steps and evaluations, were provided by Ulf Blanke (ETH Zurich, Switzerland).

4.8.1 Subtle Arm Activity Recognition based on Motion Features



This section analyzes how well different hand activities can be detected. The objective is to distinguish between activities like opening/closing a microwave, opening/closing different cupboards or pushing buttons on various devices like a laser printer, an ink jet printer or a microwave. We have already seen, that a state-of-the-art approach using inertial sensors and a representative training data set was neither able to provide sufficient results (see Section 4.3.3) nor improve the outcome of the core on-body system (see Section 4.7.7 and Section 4.7.9). Consequently, the idea is to use a trained inertial sensor system on top of a reliable spotting system, which is able to identify the type of object the user is currently interacting with. Based on recognized object interaction events, inertial sensors are only used to distinguish between feasible arm activities that belong to the object type spotted. It is clear, that by following this approach, the achieved recognition rate cannot be improved. Instead, more detailed information about the object interactions performed (underlying subtle arm actions) can be obtained. In the following this system configuration is referred to as $(BS_{SA+ROI+RFL} + IS)_{subtleAct}$.

4.8.1.1 Training and Classification

First, $BS_{SA+ROI+RFL}$, which was able to provide reasonable spotting results (EER of 79%) so far, is used to spot and to identify object interactions within a continuous data stream. Besides this, an inertial sensor system as described in Section 4.3.1[53] is used to recognize subtle arm actions on top of object interactions, that have been already spotted. Inertial systems were trained for each object with the objective to distinguish between a finite set of feasible object-related hand activities only. Each system uses a statistically significant amount of training data including all possible arm actions related to a specific object (e.g. the inertial system for "Microwave" should be able to distinguish between "open", "close", "clean device" and "start program"). To make the training procedure simpler, background activities (i.e. all activities a person can normally perform, that are not related to the object) were not considered. This means, that falsely spotted object interaction events (delivered by $BS_{SA+ROI+RFL}$) cannot be rejected and therefore the precision of the system cannot be improved. Nevertheless, the system will provide more detailed information about the object interactions performed. The integration is done as follows: For each object spotted in TSS_i, the included samples are analyzed. For each sample, scores (delivered by IS) for arm actions belonging to the object considered are compared and the highest-rated subtle arm action is taken as winner for this sample. Finally, the arm action with the highest occurrence value (based on samples) within a TSS_i is taken as subtle arm action for the related object. Consequently every object is replaced with subtle arm

[53] The "Final Segmentation Procedure" step was not applied.

actions delivered by IS. Table 4.4 (in Section 4.3.3.3) lists activities, which were considered in this section and Figure 4.58 visualizes the system concept again.

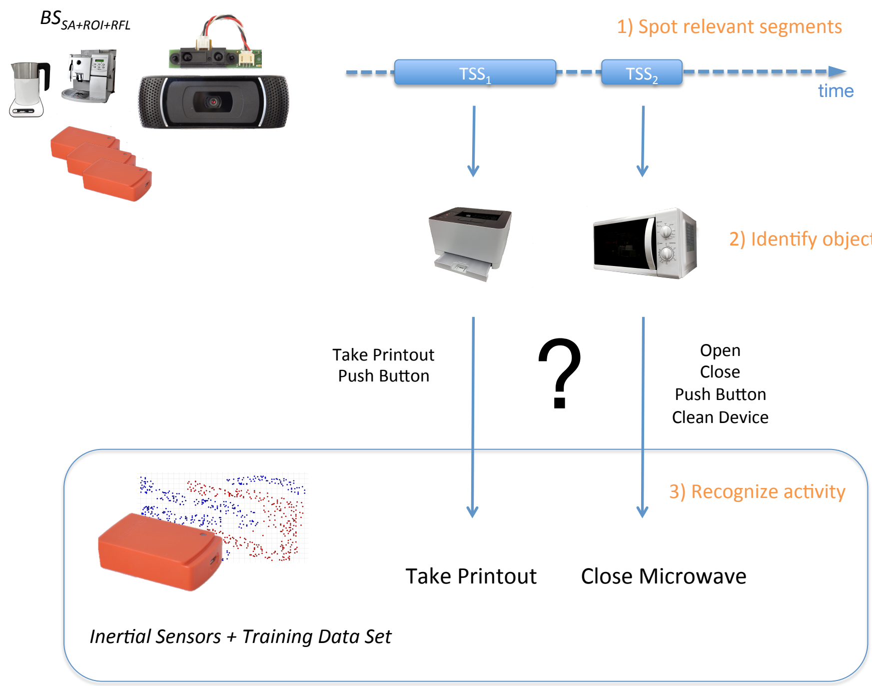


Figure 4.58: Fusion Concept: $(BS_{SA+ROI+RFL} + IS)_{subtleAct}$.

4.8.1.2 Evaluation

Figure 4.59 shows the $(1 - precision) - recall$ curve for various $thr_{svmScore}$ values. The maximum $recall$ that could be achieved was 58% with a corresponding $precision$ of 39%. The EER is 44%. This means, that compared to $BS_{SA+ROI+RFL}$ the maximum $recall$ decreased by a significant 32%, the corresponding $precision$ by 31% and the EER by 35%. As already mentioned, a recognition rate decrease was expected as the number of classes increased by 11 from 16 object interactions to 27 underlying subtle arm activities. Consequently the recognition problem became more difficult. Table 4.35 shows the results achieved for specific activities. It can be seen that the state-of-the-art inertial system even provides poor results in a scenario where subtle arm actions for specific devices have to be distinguished. For almost all devices with several underlying arm actions, the system was not able to provide reasonable results. For the devices "Battery Charger", "Climatic Control", "Shutter", "Laser Printer", "Ink Printer", "Wall Cupboard" and "Cupboard" inertial sensors do not provide enough information to solve the recognition problem although these devices include only two different underlying arm activities. This fact again illustrates the challenge of recognizing subtle arm actions with almost no characteristic movement patterns using state-of-the-art inertial sensor approaches. Only arm activities related to the "Microwave" were partly recognized. Therefore, it was possible to recognize "closing" and "cleaning" events with $recall_{act}$ values of 56% and 85%. However, the EER_{act} values for these activities were also quite low (31% and 50%). Compared to $BS_{SA+ROI+RFL}$, the average EER_{act} decreased by 39% from 74% to 35%.

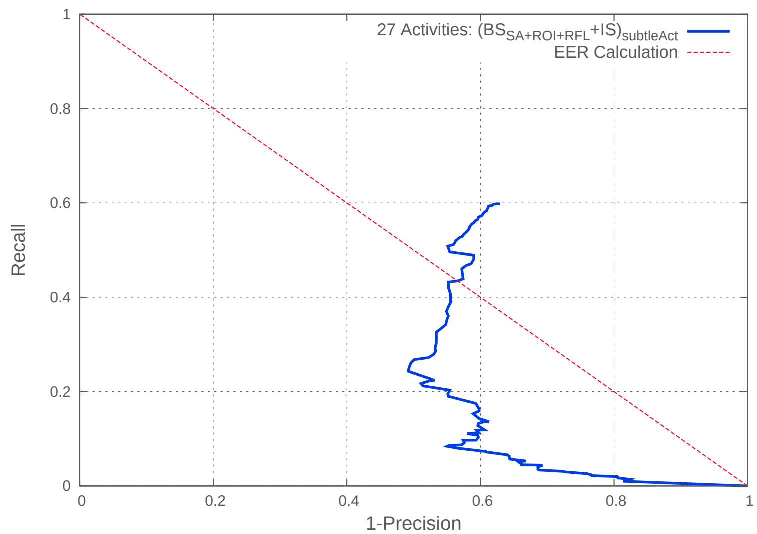
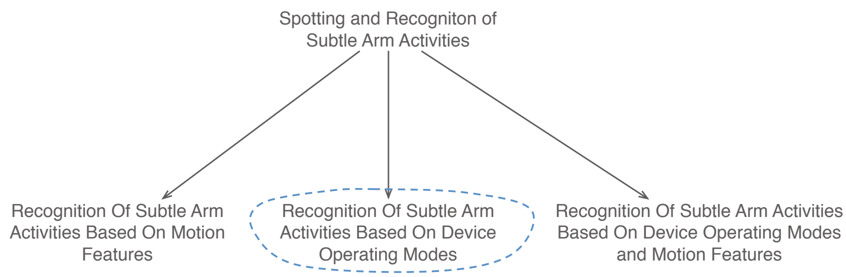


Figure 4.59: $(BS_{SA+ROI+RFL} + IS)_{subtleAct}$ (blue curve; optimized basic system + smart appliances + sub-room level location + forearm location + inertial system): $(1 - precision)$ – $recall$ curve.

Table 4.35: Activity overview for $(BS_{SA+ROI+RFL} + IS)_{subtleAct}$: $recall_{act}$, $precision_{act}$ and EER_{act}.

Activity	$recall_{act}$	$precision_{act}$	EER_{act}
Battery Charger: Put Battery In	0	0	0
Battery Charger: Remove Battery	93	26	64
Coffee Machine	89	69	85
PC	74	91	87
Air Conditioner	87	94	0
Climatic Control: Turn Left	85	36	53
Climatic Control: Turn Right	0	0	0
Shutter: Open	100	17	13
Shutter: Close	0	0	0
Use Light Button Switch	54	68	0
Microwave: Clean Device	56	24	31
Microwave: Device Closed	85	25	50
Microwave: Start Program	18	67	0
Microwave: Device Opened	4	10	0
Ethernet Connector	90	79	83
Ring Binder	93	96	95
Power Socket	70	31	59
Laser Printer: Take Printout	92	22	58
Laser Printer: Push Button	0	0	0
Ink Printer: Take Printout	93	29	37
Ink Printer: Push Button	0	0	0
Scanner	95	46	59
Wall Cupboard: Closed	23	86	0
Wall Cupboard: Opened	96	39	30
Cupboard: Closed	0	0	0
Cupboard: Opened	78	46	78
Water Tap	100	54	74
ØAverage	58	39	35

4.8.2 Subtle Arm Activity Recognition based on Device Operating Modes



As subtle arm actions aiming at object interactions normally involve operating mode/object changes (e.g. close microwave), the idea of this section is to use information coming from smart devices to recognize subtle arm activities on top of spotted object interactions (in the following referenced by $(BS_{SA+ROI+RFL} + SA)_{subtleAct}$). Section 3.5 and Section 3.4 showed how mainstream electronic devices and a water tap can be turned into smart devices using mainstream sensors and easy-to-set-up systems based on a minimal training data set. $BS_{SA+ROI+RFL}$ already includes these techniques to improve the recognition rate of the basic system. However, so far only information about the existence of operating mode changes was used to reject unused objects (objects which have not changed their operating modes). In the following, more detailed information about the new state of a device is used. The next list shows monitored devices and operating modes observed:

- Battery Charger: "Battery Charging" and "Idle".
- PC: "On" and "Off".
- Microwave: "Door Open", "Door Closed", "Heating" and "Idle".
- Coffee Machine: "Mode: Brew Espresso", "Mode: Brew Coffee" and "Idle".
- Air Conditioner: "On" and "Off".
- Scanner: "On", "Off", "Scanning" and "Idle".
- Water Tap: "Fill Small Cup ($\leq$ 500 ml)", "Fill Big Pot ($>$ 500 ml)", "Tap Turned Off".

It is clear, that operating mode changes can be mapped to arm actions as for example an operating mode change from "Door Open" to "Door Closed" must be based on the subtle hand activity "close door". Consequently, subtle arm activities shown in Table 4.36 were considered in this section. The microwave device will not change its operating mode during the cleaning procedure. When the person is cleaning the microwave inside, the microwave door must be open during that time. Therefore a "Clean Device" event is assigned to each spotted microwave event which fulfills this condition. Furthermore, the microwave door must already be open before the event starts and must also still be open after the event ends[54]. Besides, even more complex hand actions such as filling a big pot with water or filling a small cup can't be distinguished

[54]Note: For this rule the real time range of spotted events was considered without the time deviations of +/- 3 seconds.

by inertial systems as they are based on exactly the same hand activity. Another example is to recognize whether a device was turned on or off. Such activities can only be recognized by inertial sensors if the system knows the starting mode of each device and tracks device changes. An important precondition of such a system is, that all people are equipped with sensors, which makes it hard to set up in real-life environments (e.g. monitoring of dementia patients: relatives must also be equipped with sensors). $(BS_{SA+ROI+RFL} + SA)_{subtleAct}$ overcomes this problem as every device knows its' current state.

Table 4.36: Merged activity set II: 25 activities grouped by their location (kitchen, printer room, meeting room and office)

Object	Activities (Repetitions)
Microwave	Open (27), Close (27), Start (11), Clean (16)
Coffee Machine	Make Espresso (14), Make Coffee (13)
Power Socket	Connect Cable (27)
Cupboard	Use Cupboard (54)
Wall Cupboard	Use Wall Cupboard (53)
Ethernet Connector	Connect Cable (30)
Water Tap	Fill Big Cup (14), Fill Small Cup (13)
Battery Charger	Put In Empty Battery (15)
	Remove Battery While Charging (14)
Laser Printer	Use Printer (27)
Ink Printer	Use Printer (27)
Climatic Control	Use Climatic Control (27)
PC	Turn On (55)
Scanner	On (29), Off (27), Scan Document (27)
Air Conditioner	On (27), Off (27)
Light-Switch	Use Switch (55)
Ring Binder	Take-Put Binder Back (56)

4.8.2.1 Training and Classification

Figure 4.60 visualizes the system concept. $BS_{SA+ROI+RFL}$ is used to spot object interaction events. If the spotted event is related to a smart device, the new operating mode is used as underlying subtle arm action. The precondition that an operating mode change is existent is already fulfilled as $(BS_{SA+ROI+RFL}+SA)_{subtleAct}$ (or in fact $BS_{SA+ROI+RFL}$) is based on that assumption. Consequently, there is no need to add any further training or configuration steps.

4.8.2.2 Evaluation

$(BS_{SA+ROI+RFL} + SA)_{subtleAct}$ is able to reach a maximum *recall* of 92%, a corresponding *precision* of 57% and an *EER* of 69%. Figure 4.61 shows the $(1 - precision) - recall$ curve. This means, that the *recall* value of $BS_{SA+ROI+RFL}$ could be raised by 2% while increasing the number of classes from 16 to 25. However, the *precision* and the *EER* went down by 13% and 10% respectively. This fact is partly due to the relatively high amount of existing fragmentations and merges (see Section 4.7.8). So far, spotted events related to a specific object had no impact on the precision of $BS_{SA+ROI+RFL}$ due to the chosen evaluation technique. However, when splitting an object class into several more detailed classes, one object label can still be hit while creating a significant amount of insertions because of false classified fragmentations. Table 4.37 shows the evaluation results for specific activities. It can be seen that in cases of activities related to the "Air Conditioner" very high $recall_{act}$ and $precision_{act}$ values (between 85% and 96%) were achieved. Besides, EER_{act} values are also very high for those arm actions (Figure 4.62 shows that the calculated EER_{act} for "Turn Device Off" is only zero from a mathematical

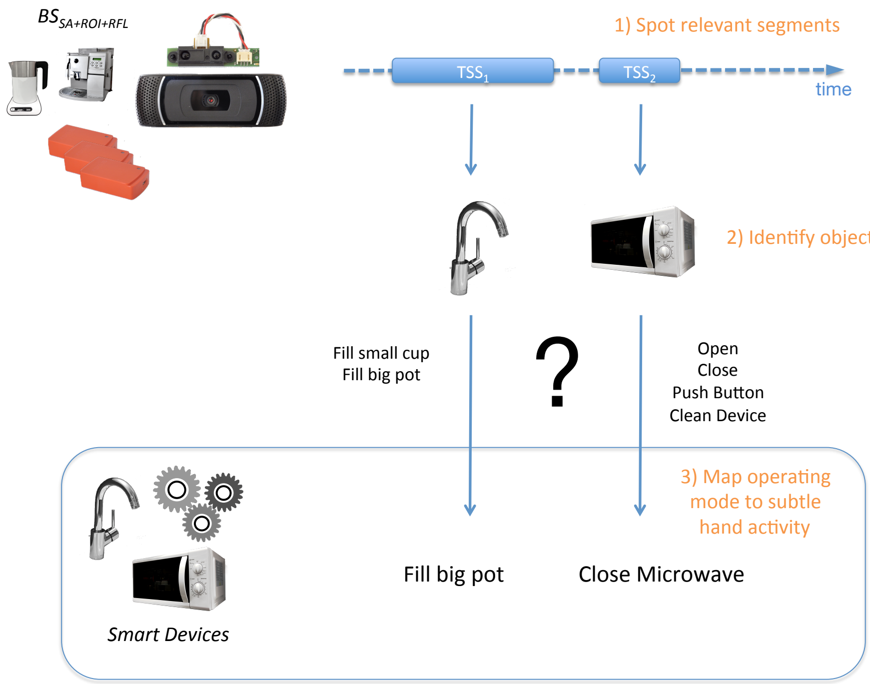


Figure 4.60: Fusion Concept: $(BS_{SA+ROI+RFL} + SA)_{subtleAct}$.

point of view). Other activities such as arm actions related to the "Microwave" can reach a high $recall_{act}$ (88% and more) but having $precision_{act}$ values between 34% and 43% only. On average the EER_{act} for activities is 69%, which is only 5% lower than was achieved by $BS_{SA+ROI+RFL}$. In return, the number of classes (activities) – and therefore the accuracy level – of the recognition system increased from 16 to 25.

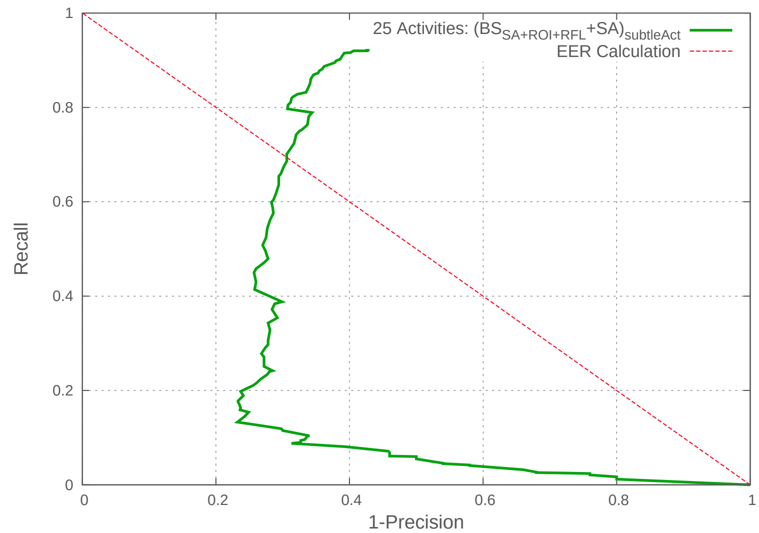


Figure 4.61: $(BS_{SA+ROI+RFL} + SA)_{subtleAct}$ (green curve; optimized basic system+smart appliances+sub-room level location+forearm location+smart appliances high-level information): $(1 - precision) - recall$ curve.

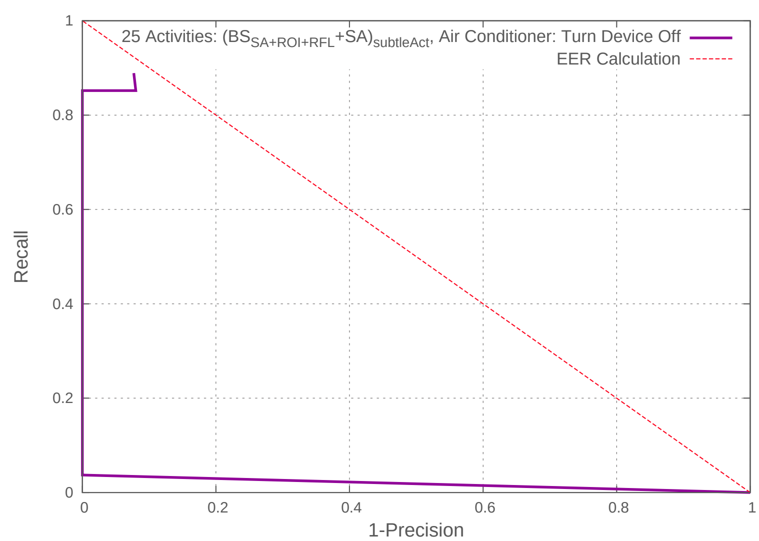
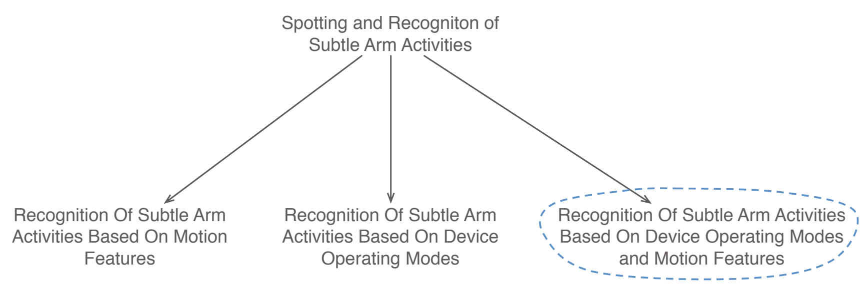


Figure 4.62: $(BS_{SA+ROI+RFL} + SA)_{subtleAct}$ (purple curve; optimized basic system+smart appliances+sub-room level location+forearm location+smart appliances high-level information): $(1 - precision) - recall$ curve for "Air Conditioner: Turn Device Off".

Table 4.37: Activity overview for $(BS_{SA+ROI+RFL} + SA)_{subtleAct}$: $recall_{act}$, $precision_{act}$ and EER_{act}.

Activity	$recall_{act}$	$precision_{act}$	EER_{act}
Battery Charger: Put Battery	87	43	47
Battery Charger: Remove Battery	86	44	64
Coffee Machine: Brew Coffee	92	71	82
Coffee Machine: Brew Espresso	79	61	79
PC	74	91	87
Climatic Control Panel	93	86	89
Air Conditioner: Turn Device Off	89	92	0
Air Conditioner: Turn Device On	85	96	93
Microwave: Clean Device	88	37	44
Microwave: Device Closed	96	43	77
Microwave: Start Program	100	38	64
Microwave: Device Opened	100	34	48
Ethernet Connector	90	79	83
Ring Binder	93	96	95
Power Socket	70	31	59
Laser Printer	96	59	83
Ink Printer	96	59	79
Light-Shutter Switch	98	79	84
Scanner: Turn Device Off	82	82	82
Scanner: Turn Device On	100	63	76
Scanner: Scan Document	93	17	22
Wall Cupboard	98	66	62
Cupboard	78	68	78
Water Tap: Fill Big Pot	100	52	79
Water Tap: Fill Small Cup	92	48	66
ØAverage	90	61	69

4.8.3 Subtle Arm Activity Recognition based on Device Operating Modes and Motion Features



So far, we have used $BS_{SA+ROI+RFL}$ in combination with IS to recognize subtle arm activities, which were performed to interact with objects. Apart from this, we also combined $BS_{SA+ROI+RFL}$ with information coming from smart devices to map operating mode changes to underlying arm actions. As both systems are based on different pre-conditions or techniques, they are able to recognize a diverse, partly overlapping activity set. Using inertial sensors, it is nearly impossible to recognize if a person has pushed a button in order to turn the device on or off (assuming that the system does not keep previous states in mind). In contrast, smart devices are able to provide this information. The idea of this chapter is to merge both approaches and to maximize the amount of recognizable subtle arm actions. In the following, the system is referred to as $(BS_{SA+ROI+RFL} + IS + SA)_{subtleAct}$.

4.8.3.1 Training and Classification

Figure 4.63 shows the system concept. The fusion concept is as follows: $BS_{SA+ROI+RFL}$ is used to spot relevant sequences and to identify the object the user is interacting with during that time. For each object the underlying subtle arm action is recognized using operating mode information in the case of a smart device or inertial sensors otherwise (if more than a single underlying arm action is possible). The same systems as introduced in the two previous sections were used. Finally, 32 subtle arm activities as shown in Table 4.38 were considered in this section.

4.8.3.2 Evaluation

Figure 4.64 shows the $(1 - precision) - recall$ curve. It can be seen, that a *recall* of 74% with a corresponding *precision* of 43% can be reached. The EER value is 51%, which is 7% higher than it was for $(BS_{SA+ROI+RFL} + IS)_{subtleAct}$, but 18% lower than reached by $(BS_{SA+ROI+RFL} + SA)_{subtleAct}$. Table 4.39 gives a detailed overview about the achieved classification rates for specific activities. There, the average object EER_{obj} decreased by 22% compared to $(BS_{SA+ROI+RFL}+SA)_{subtleAct}$, but increased by 12% compared to $(BS_{SA+ROI+RFL}+IS)_{subtleAct}$. In addition, the amount of classes (relevant activities) increased by 5 / 7 compared to $(BS_{SA+ROI+RFL} + IS)_{subtleAct}$ / $(BS_{SA+ROI+RFL} + SA)_{subtleAct}$.

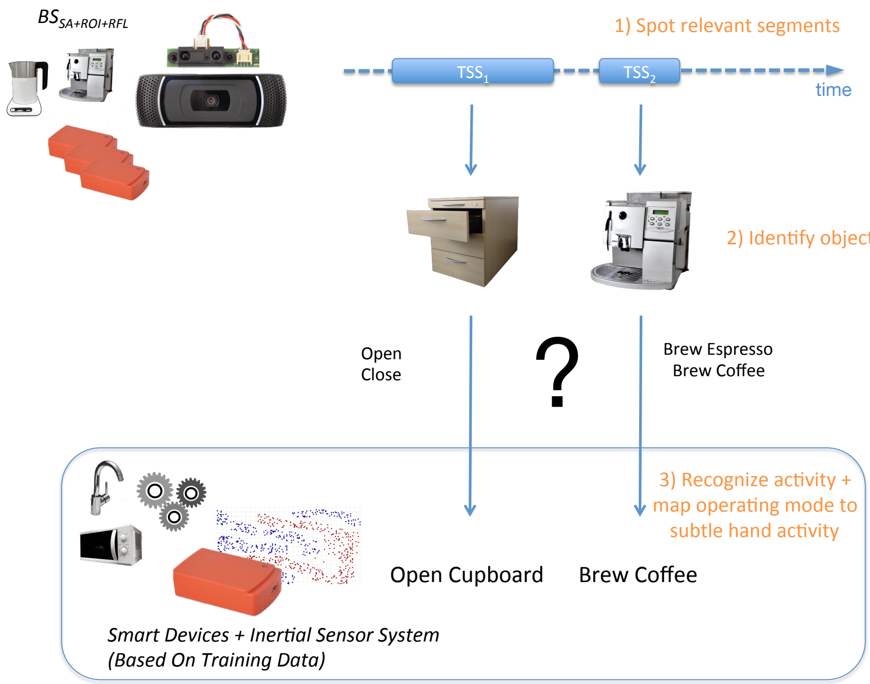


Figure 4.63: Fusion Concept: $(BS_{SA+ROI+RFL} + IS + SA)_{subtleAct}$.

Table 4.38: Merged activity set III: 32 activities grouped by their location (kitchen, printer room, meeting room and office) + fusion technique FT (SA = Smart Appliance, IS = Inertial Sensors, AM = Automatic Mapping).

Object	Activities (Repetitions)	FT
Microwave	Open (27), Close (27), Start (11), Clean (16)	SA
Coffee Machine	Make Espresso (14), Make Coffee (13)	SA
Power Socket	Connect Cable (27)	AM
Cupboard	Open (27), Close (27)	IS
Wall Cupboard	Open (27), Close (26)	IS
Ethernet Connector	Connect Cable (30)	AM
Water Tap	Fill Big Cup (14), Fill Small Cup (13)	SA
Battery Charger	Put In Empty Battery (15)	SA
	Remove Battery While Charging (14)	SA
Laser Printer	Take Printout (12), Push Button (15)	IS
Ink Printer	Take Printout (15), Push Button (12)	IS
Climatic Control	Turn Left (13), Turn Right (14)	IS
PC	Turn On (55)	AM
Scanner	On (29), Off (27), Scan Document (27)	SA
Air Conditioner	On (27), Off (27)	SA
Light-Shutter Switch	Light Button (28)	IS
Shutter Switch	Shutter Turn Left (8), Shutter Turn Right (19)	IS
Ring Binder	Take Binder Out (28), Put Binder Back (28)	IS

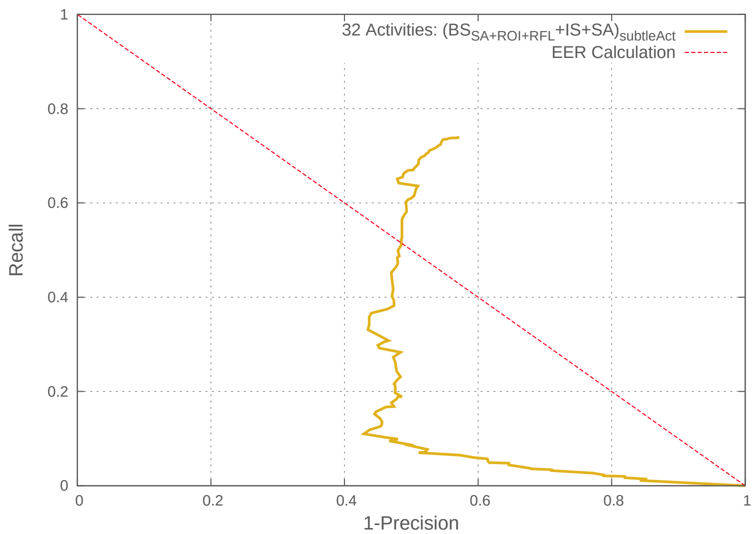


Figure 4.64: $(BS_{SA+ROI+RFL}+IS+SA)_{subtleAct}$ (orange curve; optimized basic system+smart appliances+sub-room level location+forearm location+smart appliances high-level information+inertial system): $(1 - precision) - recall$ curve.

Table 4.39: Activity overview for $(BS_{SA+ROI+RFL} + IS + SA)_{subtleAct}$: $recall_{act}$, $precision_{act}$ and EER_{act}.

Activity	$recall_{act}$	$precision_{act}$	EER_{act}
Battery Charger: Put Battery In	87	43	47
Battery Charger: Remove Battery	86	44	64
Coffee Machine: Brew Coffee	92	71	82
Coffee Machine: Brew Espresso	79	61	79
PC	74	91	87
Climatic Control: Turn Left	85	36	53
Climatic Control: Turn Right	0	0	0
Air Conditioner: Turn Device Off	89	92	0
Air Conditioner: Turn Device On	85	96	93
Shutter: Open	100	17	13
Shutter: Close	0	0	0
Press Light Button Switch	54	68	0
Microwave: Clean Device	88	37	44
Microwave: Device Closed	96	43	77
Microwave: Start Program	100	38	64
Microwave: Device Opened	100	34	48
Ethernet Connector	90	79	83
Ring Binder	93	96	95
Power Socket	70	31	59
Laser Printer: Take Printout	92	22	58
Laser Printer: Push Button	0	0	0
Ink Printer: Take Printout	93	29	37
Ink Printer: Push Button	0	0	0
Scanner: Turn Device Off	82	82	82
Scanner: Turn Device On	100	63	76
Scanner: Scan Document	93	17	22
Wall Cupboard: Closed	23	86	0
Wall Cupboard: Opened	96	39	30
Cupboard: Closed	0	0	0
Cupboard: Opened	78	46	78
Water Tap: Fill Big Pot	100	52	79
Water Tap: Fill Small Cup	92	48	66
ØAverage	72	46	47

4.9 Results and Summary

This chapter focused on various sensor fusion techniques based on approaches ranging from simple, model-based approaches to systems using statistically significant, large training data sets. The objective was to solve the introduced problem of spotting object interactions coming from subtle hand actions within a continuous data stream containing daily life activities. We saw that a state-of-the-art technique based on inertial sensors and statistically relevant training data is not able to provide reasonable results (EER of 18%).

A core on-body system (BS) based on simple configurations and a minimal set of training data which was able to outperform the motion system (EER improvement of 29%) was introduced. Several model-based approaches and systems using minimal data sets in combination with the core on-body system were evaluated. The idea was to restrict the search space of the problem considered by sensor fusion and to compensate in this way the fact that only minimal training data and easy to perform one-time measurements and configurations were used. This way, an improvement in terms of *recall*, *precision* and EER was achieved. While an EER of 79% is not perfect, it is sufficient for many real-life applications. Apart from that, the system is a good starting point for further sensor fusion approaches. Finally, it was also shown, that fusing the inertial sensor system and the core on-body system could not improve the recognition quality at all. Also, after replacing the wearable camera sensor of the core on-body system with the inertial system, the results achieved could not be maintained (maximum EER of 27%).

Table 4.40 summarizes the fusion approaches evaluated and the corresponding results again. The best results are still achieved using $BS_{SA+ROI+RFL}$ (EER of 79%; core system + operating mode detection + sub-room level activity monitoring + magnetic signatures of objects) assuming the existence of smart appliances, $BS_{ROI+RFL}$ (EER of 68%; core system+sub-room level activity monitoring + magnetic signatures of objects) when minimal infrastructure is in focus and $BS_{TF+MoL+RFL}$ (EER of 55%; core system + time features + standing detection + magnetic signatures of objects) when infrastructure instrumentation is not possible. A more precise event-based evaluation approach could improve this fact. However, when using a time frame-based evaluation technique it could be seen that the number of insertions is insignificantly small compared to the amount of available time frames. Consequently, almost all time samples included in the experiments were recognized correctly. However, the amount of insertions is not inconsiderably low compared to the number of activities performed.

Besides that, it was shown that the introduced sensor fusion approaches could significantly reduce the number of classification steps as well as the amount of analyzed images. This fact results in a meaningful performance improvement. However, some fusion approaches have not been able to increase *recall*, *precision* and EER (namely BS_{MoL}[55], BS_{HM}[56] and BS_{TF}[57]), they are still able to contribute to a better system performance and a reduced processing time.

Next, this work went a step further and analyzed how well subtle arm actions performed when interacting with objects, that can be recognized. Depending on the sensor modalities used, three different activity sets containing 25, 27 and 32 arm actions were considered. It was shown, that a state-of-the-art approach based on inertial sensors was unable to provide sufficient results (EER of 13%). Consequently, $BS_{SA+ROI+RFL}$[58] was used as baseline for further processing steps. On top of spotted object interaction events, the inertial system was used to distinguish between different object-related arm actions (e.g.: microwave opened, closed, cleaned or button on device pushed). Besides inertial sensors, the impact of using information about operating mode changes coming from smart devices was investigated. Operating mode changes and the resulting new operating modes were mapped to underlying arm actions (e.g. "new microwave mode: door open" was mapped to the human activity "microwave opened"). Both approaches provide different activity sets. For example using only inertial sensors it is nearly impossible to distinguish between turning a device on or off as both activities are based on the same arm

[55] Core system + standing detection
[56] Core system + hand movement intensity
[57] Core system + time features
[58] Core system + operating mode detection + sub-room level activity monitoring + magnetic signatures of objects

movement (pushing the On-Off button). Smart devices however, can deliver such information with ease. Hence, both techniques were merged in order to get a maximized activity set. Figure 4.65 shows again the achieved $(1-precision)-recall$ curves for all evaluated systems. It can be seen, that the best results are achieved by using $BS_{SA+ROI+RFL}$ (core system + operating mode detection + sub-room level activity monitoring + magnetic signatures of objects) in combination with information coming from smart devices. Here, 27 activities can be distinguished with an EER of 69%. The combination of smart devices and inertial sensors is able to maximize the amount of recognizable activities (32 activities), but provides an EER which is 18% lower. The combination of $BS_{SA+ROI+RFL}$ (core system + operating mode detection + sub-room level activity monitoring + magnetic signatures of objects) with inertial sensors or even a system using inertial sensors only delivers the worst results (EER of 44% and 13%). This fact shows again, that the introduced systems can significantly improve the recognition rates of the type of activity spotting problem considered.

Table 4.40: Overall system comparison: $recall$, $precision$, EER and percentage reduction of classification steps (CS_R) as well as analyzed images (AI_R) compared to $BS_{optDistOH}$ (optimized basic system).

System	$recall$	$precision$	EER	CS_R	AI_R
IS	80	3	18	–	–
IS'_{HH}	32	3	0	–	–
$IS'_{PROX+HH}$	42	9	27	–	–
$BS_{optDistOH}$	75	22	47	–	–
BS_{BT}	82	31	56	61	4
BS_{ROI}	90	40	61	80	28
BS_{RFL}	78	29	55	66	19
BS_{MoL}	75	22	47	3	3
BS_{HM}	75	22	46	35	42
BS_{TF}	75	22	47	14	21
BS_{SA}	81	35	53	51	0
BS_{BT+ROI}	87	48	63	84	38
BS_{BT+RFL}	82	46	64	85	36
$BS_{ROI+RFL}$	87	56	68	90	49
$BS_{BT+ROI+RFL}$	84	60	69	91	53
$BS_{TF+MoL+RFL}$	78	30	55	70	32
$BS_{TF+MoL+ROI+RFL}$	86	57	68	91	56
$BS_{SA+ROI+RFL}$	90	70	79	94	63
$BS_{TF+MoL+SA+ROI+RFL}$	89	70	79	95	68
$BS_{SA+ROI+RFL}+IS$	89	70	79	94	63
$BS_{ROI+RFL}+IS$	85	56	68	90	49
$BS_{TF+MoL+RFL}+IS$	75	30	55	70	32

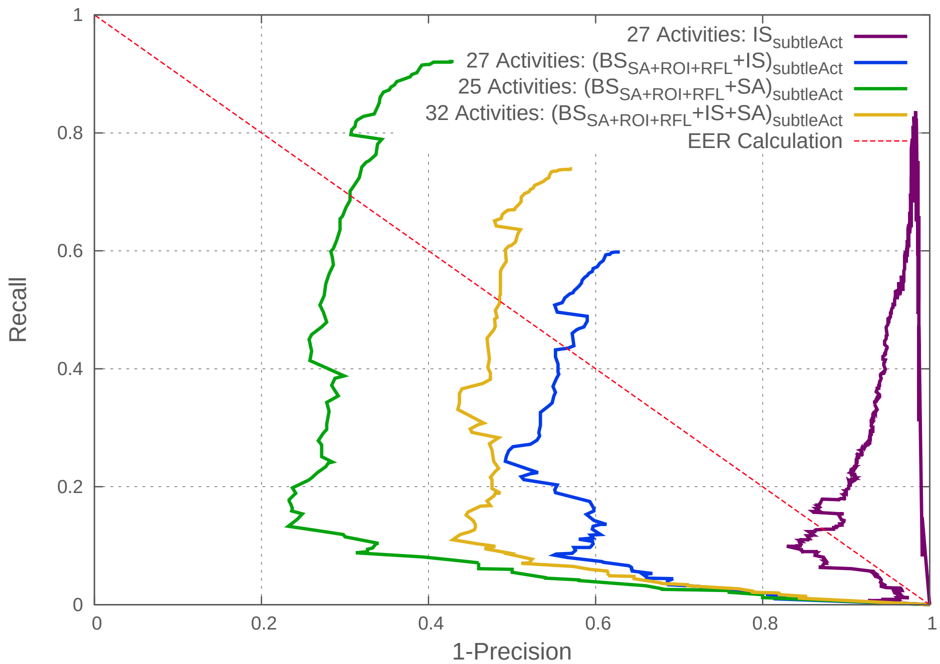


Figure 4.65: $(1 - precision) - recall$ curves (based on $thr_{svmScore}$) for $(BS_{SA+ROI+RFL} + IS + SA)_{subtleAct}$ (orange curve; optimized basic system+smart appliances+sub-room level location+forearm location+inertial system+smart appliances high-level information), $(BS_{SA+ROI+RFL} + SA)_{subtleAct}$ (green curve; optimized basic system+smart appliances+sub-room level location+forearm location+smart appliances high-level information), $(BS_{SA+ROI+RFL} + IS)_{subtleAct}$ (blue curve; optimized basic system+smart appliances+sub-room level location+forearm location+inertial system), $IS_{subtleAct}$ (purple curve; inertial system).

4.10 Conclusion

In summary, the following results were achieved:

- We have seen, that state-of-the-art motion-based approaches, which have been widely and successfully used by the research community for several activity spotting problems, are neither able to deliver sufficient results with respect to subtle, barely distinguishable arm activities nor to contribute to a performance improvement of the core on-body system.
- The proposed wearable on-body system achieves a reasonable performance in terms of spotting object interactions despite the fact that it is configured by simple-to-perform one-time measurements (including a one-shot object capturing) and configurations only. Moreover, the system does not rely on environmental instrumentation.
- The core on-body system is based on a state-of-the-art computer vision object recognition approach but combined with a one-shot training procedure. The lack of training images is compensated by fusing the system with models related to common activity pre-conditions. Several fusion approaches with the core on-body system were considered to restrict the search space further and improve recognition rates achieved so far.
- It was shown, that fusing the core on-body system with easy to deploy (considered systems can be even installed by people with limited technical knowledge) location-based approaches as well as device operating mode information results in a significant performance improvement.
- In contrast, motion features like a person's basic modes of locomotion or hand movement intensities could not contribute to a performance improvement although related concepts have worked well for similar spotting problems. This fact shows again, that subtle arm activities are quite similar to motions regularly performed during other activities and therefore systems, that are only based on motion features, are not able to achieve sufficient results. However, these systems as well as time related features were able to significantly reduce the amount of processed images and classification steps. Hence they may contribute to a significant runtime improvement.
- Fusing several of the proposed systems results in immense improvements in performance. This work highlighted three fusion approaches, which were adapted to various real-life conditions. In-depth evaluations confirmed the results achieved so far.
- Subtle arm activities have been recognized based on identified object interactions and a motion system as well as device operating modes. Again, the motion system delivers significantly worse results than using information about device use-modes. The combination of both approaches maximizes the amount of recognizable classes but is not able to reach a similar accuracy when considering device use-modes only.

Finally, this work shows, that a difficult spotting problem can be solved by using a multi-modal sensor approach based on:

- Models, that are configured by simple and easy to perform one-time measurements. Hence they do not rely on statistically significant, large training data sets.
- Systems, that rely on minimal training data sets, that could be collected with less effort (one-time location walk-through) even by technical laymen.
- Systems, that can be integrated in existing environments unobtrusively.
- Low-cost sensors and therefore affordable systems.

However, the use of the system proposed in common real-life scenarios may be restricted by the following aspects:

- First, due to the fact, that the vertical distance between objects and the user's hand is a core information of the system, it can hardly be used in scenarios, in which objects are frequently relocated (e.g. forks or plates). There, an unacceptable re-configuration effort would have to be performed to update the vertical position of objects.

- Secondly, as the core on-body system recognizes objects based on a wrist-mounted camera, its use may be restricted due to computer vision-related problems. For example, issues related to computation time as well as light conditions can cause serious problems in common real-life applications.

- Thirdly, the best EER of 79% can only be achieved if information about device operating modes is available. Many objects can hardly be equipped with sensors in an unobtrusive way in real-life environments in order to be able to provide information about their use-modes. We have seen, that the best EER that could be achieved without such device information is 68%, which might be too low for many applications.

The last aspect leads to the following question:

What other activity recognition systems / methods may be useful to improve the performance of the core on-body system?

As was discussed in this section, a large amount of concepts and methods exist that may contribute to the problem observed. In my opinion a promising approach that could be used in combination with the proposed system is described in [CBL12]. There, a capacity sensor has been unobtrusively integrated into a common wristband. Based on a statistically significant amount of training data (several repetitions for each hand gesture), the system was able to differentiate between a finite set of hand activities. However, this fact would destroy the idea of minimal training data. Consequently, the idea would be to use such a system to indicate common valid activity pre-conditions in order to further reduce the search space of the on-body system. In contrast to inertial systems aiming to detect specific hand motions, such systems cover a different aspect, that could be very useful for the problem of spotting subtle hand activities: Features describing processes that happen under the skin during hand activities. Such features may be very useful in the case of subtle, barely distinguishable arm activities, that cannot be reliably described by motion features.

Besides that, other methods and algorithms can be investigated such as:

- Improving the computer vision-based object recognition method by replacing the linear SVM kernel with an RBF or histogram intersection kernel (see [BOV03]) or by using Joint Boosting approaches (e.g. [MF09]).

- Defining dependencies between activities and object interactions using context-free grammars (e.g. [RA06]).

- Improving the inertial system by using longest common subsequence (e.g. [NDRCT12]) or derivative dynamic time warping (e.g. [BD13]) methods.

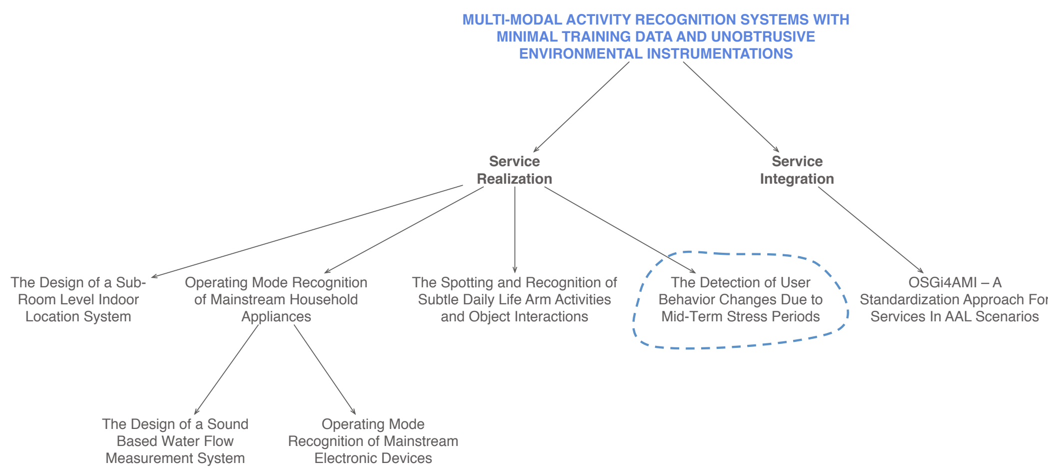
5

Detection of User Behavioral Changes Due to Medium-Term Stress Periods

This chapter is based on my work published in

Gerald Bauer and Paul Lukowicz. *Can Smartphones Detect Stress-Related Changes in the Behavior of Individuals?*, in 'PerCom Workshops', IEEE, 2012, pp. 423-426

and supervised work shown in [Fin11] and [Bin11].



5.1 Introduction

So far the thesis introduced various approaches based on minimal training effort and designed for large-scale and real-world scenarios within closed environments. This section goes a step further and introduces a behavior monitoring system that is completely free from any environmental instrumentation. Behavior monitoring systems can be designed for different purposes such as health or social life monitoring. This work focuses on behavioral changes caused by a widespread disease, which has gained immense importance during recent years: Stress and burnout syndromes. As will be shown in Section 5.2 many approaches have already addressed this issue. However, to a large part such systems are mainly based on physiological parameters derived from heart rate, galvanic skin response or voice analysis. Although the necessary sensing systems are becoming more and more wearable and usable in real-life scenarios, they are still not really accepted by society and are rated as too obtrusive. One of the reasons is, that sensors have to be attached directly to the skin and hence people may feel uncomfortable while wearing them. Besides, it is clear, that stressed people become even more negatively affected by wearing strange on-body sensors, especially if they have to worry about battery supply or sensor displacements.

This section introduces a novel stress detection system that works towards the solution of these issues. The proposed system is based on smartphone integrated sensors only. The fact, that people like to carry their mobile phones around all day makes such devices unobtrusive and powerful monitoring systems. In contrast to many previous approaches that aim at the detection of short-term stress situations, the proposed system focuses on medium-term monitoring and the detection of time periods in which users are continuously under stress. The system determines behavioral parameters on the basis of smartphone sensors. Derived behavior patterns are compared between stressful and stress-free time periods in order to detect significant behavioral changes.

A 4-7-24 (four week, seven days a week, 24 hours a day) data recording was performed with multiple students. The period observed contains about two weeks of continuous stress (examination period) and two stress-free weeks (outside the examination period). Although the scale of the experiment is small and it remains to be seen how the achieved results can be transferred to other stress situations, this section will illustrate, that the proposed approach is able to detect clear behavior changes for all participants from one period to the other.

This work was supported by the EU project INTERSTRESS, FP7-24768.

5.2 Related Work

Many research approaches have already focused on the automatic detection of stress situations. The most common way to monitor a person's stress level is to use physiological parameters (see [dSSSACdP11] [CALMT11] [CAM+11] [BAF+91] [dSSAGC+10] [ADSP11] [SCB11] [CAM+11]). There, features such as heart rate and galvanic skin response were considered. Other approaches are based on the analysis of voice samples (see [HLMA09] [ZMTA09] [RHM+02]). Such approaches showed very promising results. However the fact that sensors are attached to the skin or microphones are continuously analyzing voices of people makes them too obtrusive for real-life applications, where people have to be monitored 24 hours, 7 days a week. Because smartphones are known as unobtrusive sensor systems, they have been considered in many approaches related to behavior monitoring in general as well as for stress detection. In [CSV+12] a mobile phone-based self-report questionnaire is combined with physiological measurements using ECG (electrocardiogram), EMG (electromyography) and RSP (chest respiration). The focus was to extend the capabilities of standard self-report questionnaires which are used to detect stress situations as shown in [JPS+13]. The system shown in [GPT+] was used to investigate the relation between the user's psychological, physiological and activity information. Besides, acceleration data were used to calculate the activity intensity of the user and to combine this information with spectral power in low (LF) and high frequency (HF) bands of HRV (heart rate variability). As the ratio LF/HF is affected by strong physical activities, only the ECG signals during low activity time ranges were analyzed. In contrast to these approaches the work described in this thesis does not refer to physiological parameters. Instead, smartphones are used to determine user behavior patterns and to detect stress-related behavioral changes. The idea of using mobile phones as sensor systems in behavioral monitoring application has already been used in several approaches. In [API+11] [EPL09] [ZDC11] systems are shown that use smartphones to recognize social interactions, to infer the structure of friendship networks or human relationships as well as to analyze user behaviors in social networks. In contrast to this work, these approaches focus on the detection of behavioral changes in general and are not related to medium-term stress situations.

5.3 Research Questions and Contributions

This chapter considers the problem of stress detection in medium-term[59], real-life scenarios. So far, related approaches are largely based on physiological parameters and aim to detect single stress situations in real-time (see [ADSP11] [SCB11] [dSSSACdP11]). As sensors have to be directly attached to the skin and have to be permanently worn in order to monitor parameters like heart rate, the resulting problem is obvious: The useage of such systems is quite limited in real-world applications where people have to be monitored 24 hours a day for several weeks.

Consequently, the ***key problem*** can be described by:

How can a wearable stress detection system be designed unobtrusively and without using skin attached sensors?

A common approach is to analyze the user's voice. Amongst others, such approaches were presented in [HLMA09] [ZMTA09] [RHM+02]. However, due to problems already mentioned (see Section 5.2) they were not considered in this work.

The key idea of the proposed system is to derive user behavior patterns, that are able to reflect stress situations, from smartphone integrated sensor systems. This concept was motivated by the following aspects:

- First, approaches like [API+11] [EPL09] [ZDC11] have already shown, that smartphones can be used to derive social interactions, friendship networks and human relationships. Consequently, such devices are able to derive daily-life behavior patterns.
- Secondly, smartphones are indisputably one of the most unobtrusive sensor systems by nature. People are used to carry their mobile phones with them all day. Consequently, they provide a perfect solution for an unobtrusive 24 hour behavior monitoring problem as is considered in this chapter.

These considerations lead to the ***first research question***:

What kind of behavioral parameters, that are affected by medium-term stress situations, can be derived by smartphone integrated sensors?

In general, the following user behavior patterns, which have already been addressed by several approaches, may be affected by medium-term stress situations:

- Location Behavior: Most obviously, people facing continuous stress situations may change usual daily-life routines. For example, people may stay longer at work, they can't visit their favorite bars or they have to regularly visit new locations (e.g. hospital). Approaches related to this topic were introduced for example in [AS02] [JZZC12] [HCL+05] [KWSB04]. Consequently, behavioral patterns related to the user's location are defined in Section 5.5.1.
- Social Interaction: During stress situations, people may have less time to visit their friends or their families. Apart from that, they may have less time to attend social events (e.g. shopping malls, cinemas or crowded places in general). Consequently, they are more isolated than usual. Approaches related to social sensing systems and crowd estimates can be found for example in [EPL09] [API+11] [JP11] [WFR+12]. Behavioral parameters related to this topic are defined in Section 5.5.2.

[59]In this context, the term "medium-term" describes a time period of several weeks.

- Phone Behavior: Phone usage behavior changes may also be related to stress situations. The idea is, that people will change their phone call and SMS behavior as they have to compensate for the lack of physical meetings or they have to communicate more with colleagues in the case of job-related stress situations. The analysis of human behavior based on mobile phone usage was shown for example in [ZDC11] [BB09]. Phone call behavior features, which were used in this work, are defined in Section 5.5.3.

So far, behavioral parameters have been shown, which could be influenced by stress situations. This fact leads to the ***second research question***:

How can the impact of medium-term stress situations on defined behavioral parameters be evaluated reliably?

Obviously, it is neither easy nor ethically desirable to expose people to medium-term periods of artificially generated stress. Apart from that, recruiting people, who could eventually be under continuous stress situations is not realistic. Section 5.6 will introduce how a real-life data set, in which people were relaxed and faced medium-time periods of continuous stress to equal parts, has been recorded. The data set consists of smartphone sensor data collected by six participants for four weeks, seven days a week and almost 24 hours a day.

Finally, the ***third*** and last ***research question***, which is addressed in this chapter, is:

How well can defined behavioral parameters indicate medium-term stress periods?

Section 5.7 will introduce a detailed evaluation based on defined behavioral parameters and the recorded data set. Therefore, behavioral features of each user were derived from both time periods and compared with each other. It will be shown, that although introduced features and their interference with medium-term stress situations are user dependent, behavioral changes between nearly 40% and almost 60% can be derived.

In summary, the ***key contributions*** of this chapter were:

- The definition and evaluation (on a real life data set) of behavioral parameters (derived from smartphone integrated sensors) that are influenced by medium-term stress situations.
- The recording of a real-life, multiple user data set, which consists of a medium-term stressful and a stress-free period.
- A concept, which shows that medium-term stress situations in real-life situations can even be detected by an unobtrusive, privacy-protective approach (sensor data remain on the mobile phone), that does not rely on skin attached sensors and voice analysis.

Next, the system concept is explained in detail, followed by the introduction of behavioral parameters, the recorded data set and evaluation results.

5.4 System Concept

Todays smartphones are already equipped with many different sensor systems as shown in Figure 5.1. Examples are GPS, Bluetooth and acceleration sensors or microphones. The fact that people like to carry their smartphones around with them all day makes them a valuable and very unobtrusive sensor system for behavior monitoring applications. This work focuses on the detection of behavioral changes due to medium-term stress periods. The assumption was that people facing continuous stress situations for a medium-term time period will change their normal day-to-day behavior. A person's behavior is described by eight features related to location traces, the grade of physical social interaction and mobile based communication. Features were calculated based on real-life data recorded by mobile phones of several people during a period

Figure 5.1: Smartphone and integrated sensor systems. Many mainstream smartphones already include sensor systems such as GPS, acceleration sensors or cameras. Based on data from such sensor modalities, valuable information about the carrier can be derived.

of two weeks of continuous stress and two stress-free weeks. Consequently, collected data was grouped into two parts containing the behavior of people during the stressful period and the time afterwards, marked as normal behavior. Behavioral parameters were calculated offline and resulting patterns have been compared between the two groups.

During the data recording smartphone functionalities were not affected at all. Although GPS + WiFi positioning were enabled all the time and these sensors are known as very power consuming, the mobile phone could still be operated for more than eight hours due to the chosen sampling frequency. In this way data was gathered that cover complete daily routines. As a consequence detailed behavioral parameters could be determined.

It is worth noting, that the system does not look at specific locations or specific social contacts. These features may be more dependent on a concrete stress situation then on general stress. In contrast, abstract parameters such as "variability of contacts" or "number of locations visited" were considered as these features may be more generalizable. Evaluations show that a clear behavioral change can be seen even in such abstract parameters.

In the following, behavioral parameters, the recorded data set and the relation of defined parameters to medium-term stress periods are explained and evaluated.

5.5 Behavioral Parameters

This section introduces features observed which are used to describe a person's behavior and detect behavioral changes due to medium-term stress periods. All in all three behavioral parameter groups were considered. In the following each group and its assumed relation to stress situations is explained in detail.

5.5.1 Location Behavior

The first group is related to a person's location. Each person has some favorite places which he/she visits more often and regularly. Examples are the person's home, workplace, a favorite cocktail bar or friend's homes. From now on we will call such locations "Regions of Interests" (ROI). Each person shows specific behavior patterns related to his/her set of ROIs. Consequently, for example such patterns include the time the person usually leaves home for work, the normal working and lunch hours, regular workout times and gym visits or regularly visits of relatives or friends. The idea was that people will significantly change their location behavior during medium-term stress periods. For example, it is assumed that people will stay at work longer during stressful times and hence they will skip regular free-time activities such as jogging, gym workouts or meeting up with friends in their favorite bars. The person's location was determined using smartphone integrated GPS and WiFi sensors. As the process of GPS positioning is very power consuming and one objective was that the mobile phone's battery should last for nearly a day, GPS positioning was performed only every 10 minutes. Of course this destroys the vision of accurate user traces. However, having GPS fixes every 10 minutes is more than enough for the task in focus because significant deviations from normal retention times should be detected only. The process of finding ROIs is based on raw GPS and WiFi positioning data. Every 15 seconds a WiFi scan was performed as the scanning procedure is much less power consuming than GPS positioning. If GPS was not available but already known WiFi routers were detected, virtual positioning coordinates were used to determine the current user location. This procedure is based on the assignment of GPS fixes to WiFi routers. Hence every time a GPS fix was available, virtual coordinates were created for each discoverable WiFi router (within a specific time range). Every time a GPS fix was not available but WiFi access points with virtual GPS coordinates were present, these coordinates were used as the current location. A k-means clustering was performed on the recorded GPS coordinates to find ROIs. There, only clusters with a maximum diameter of 300 meters and in which people continuously spend more than 10 minutes were chosen. The detailed approach description is shown in [Fin11]. Figure 5.2[60] shows determined ROIs for one of the participants in Passau, Germany.

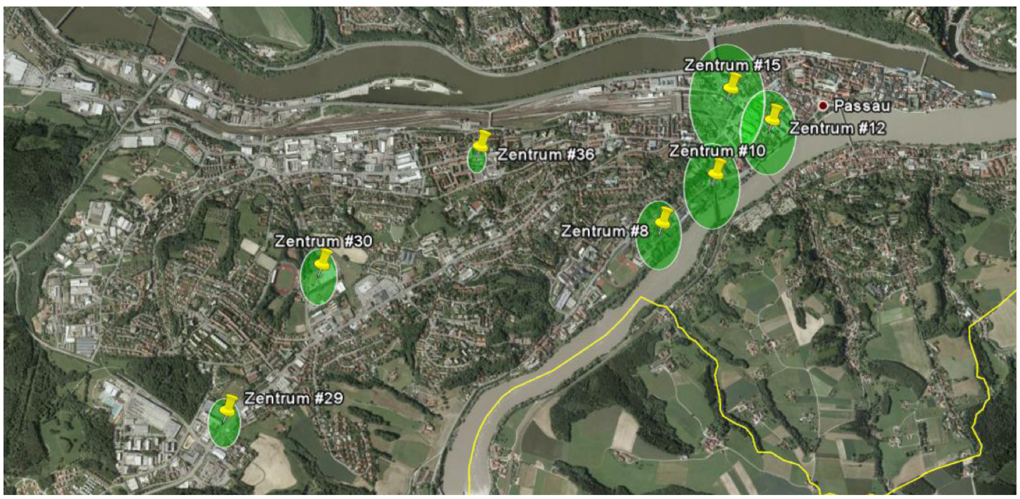


Figure 5.2: Detected ROIs (Source: [Fin11]). Seven different ROIs ("Zentrum") were found within the city of Passau for this specific person. ROIs include the city center mall, the university, the sports center and a home improvement store to name just a few.

The following behavioral features were defined based on ROI findings.

- **ROI1: Amount of important ROIs.** The overall objective was to see if people visited different locations during stressful and stress-free time periods. Therefore the amount of important ROIs were calculated. In this context, a ROI is important if it was visited at least twice as often within one time period as during the other period. In order to filter out locations that were visited often but only for a short time range (e.g. bus station, etc.), each considered ROI has to be visited for at least 20 minutes on average. Finally, ROI1 shows the amount of important ROIs in percent.

- **ROI2: Retention time.** This feature focuses on people's retention time within a ROI. The idea is, that people spent less time at locations such as gyms or their homes during medium-term stress periods. Thus it was calculated how much time people spent within ROIs during one period compared to the other. In this context only ROIs were considered in which the average retention time during one period was at least twice as long when

[60]This Figure was created by Alexander Findeis (University of Passau, Germany) and was taken from my supervised work described in [Fin11].

compared to the other. In order to filter out ROIs in which people stayed for a short time, only ROIs were considered in which people spent at least 20 minutes on average. Finally, ROI2 shows the average, absolute time deviation for a participant's ROIs in percent.

5.5.2 Social Interaction Behavior

A common known fact is that people who suffer from depression or burn-out syndrome reduce their social life activities. Thus information about how much time someone spent with other people or how long he stayed in crowded environments like public coffee houses or shopping streets, could be very useful in order to estimate the current stress level. In the case of work-related stress situations, people may spend their weekends at work and consequently, they will miss free-time activities like going to the cinema or spending time with friends. To sum up, this feature is based on the assumption that during medium-term stress periods people will significantly reduce the time they usually spend with other people or in crowded places. In contrast to location patterns, this feature does not focus on specific places like shopping malls or cinemas but how long people spent time with specific persons or how long they stayed in unspecific locations with many people around them. In the following such features are referred to as "grade of social interaction". To measure the grade of social interaction, a Bluetooth based approach was chosen. A smartphone's Bluetooth sensor was used to scan continuously (using a scan interval of 10 seconds) for available Bluetooth devices. As discovered devices are located in the near vicinity of the user and these devices belong in almost all cases to other people, estimates about the crowd density can be performed. This principle was already used in other research approaches like [NK07a] [JP11]. The following features were considered based on the Bluetooth devices discovered:

- **SI1: Overall amount of social interaction.** This feature was used to show the amount of social life activity in general. Bluetooth devices discovered within both time periods were used to analyze how the user changed his social life behavior. It is expected, that people would reduce their social life behavior significantly during stressful periods and hence are more isolated than during stress-free times.
- **SI2: Time spent with specific people.** This feature describes how much time someone has spent with another "important" person. In this context, a person is called important if he spent continuously more than 10 minutes in the immediate vicinity of the participant.

5.5.3 Phone Behavior (Calls + SMS)

The last group considers behavior patterns related to mobile communications. Usually people arrange their contacts in relationship groups. The most obvious groups are spouse, close family, friends, work colleagues or casual friends. The idea is, that people may contact specific people more frequently during stress-periods than during stress-free times. Examples could be work colleagues, project partners or even family members and friends as they want to compensate the lack of physical meetings. Consequently, anonymized information about phoning and SMS received were analyzed. Due to software issues a reliable logging of SMS sent was not possible at the time the experiment took place. Therefore, only SMS that were received were taken into account. However, as people usually reply to a SMS they receive, this information can also be used as a social interaction indicator. Based on this information the following mobile communication features were defined:

- **C1: Overall amount of calls.** This feature reflects how much a person changed his overall phone behavior. This means that the amount of phone calls performed (incoming as well as outgoing) were counted and deviations between stress-free and stressful periods were considered.
- **C2: Phone call contact behavior.** Stressed persons may also contact different people (e.g. work colleagues or friends). Consequently, this feature considers the amount of people that have been contacted either during stress-free or stressful periods in percent.

- **S1: Overall amount of SMS received.** Defined as C1 but related to SMS received.
- **S2: SMS contact behavior.** Defined as C2 but related to SMS received.

5.6 Data Set

One of the main problems was to gather a reliable data set, which can be used to evaluate behavioral parameters. It is obvious, that it is neither easy nor ethically desirable to expose people to medium-term periods of artificially generated stress. Even recruiting random people and hoping that they will be exposed to continuous stress periods for longer time periods is not realistic. As a consequence, the experiment was based on the assumption that students have to face continuous stress situations during examination periods and that they are quite relaxed afterwards. Hence seven students were monitored for four weeks, 24 hours a day. In order to gather equal amounts of data for both periods, the data recording last two weeks during the examination period and two weeks after. Only students were chosen, who rated their exams as very important but had almost no reasonable chance of passing them. In this way, a significant stress level could be assumed.

Every student was equipped with an Android-based smartphone (Google Nexus One or HTC Desire), with a sensor data logger application running in the background[61]. Raw sensor data from the following sensors was stored anonymously on the phone's SD card: GPS, WiFi, Bluetooth, Accelerometer, microphone, mobile web usage, phone calls made and SMS received. As students had to replace their own mobile with the provided smartphone during the experiment, the gathering of real-life data could be guaranteed. The received data set consists of more than 40 GB.

An important outcome of the data recording performed was, that students confirmed that they were not affected by the system in their daily-life. This fact shows, that behavioral parameters can be calculated in a very unobtrusive and a socially accepted way by using smartphones. A follow-up data recording was planned where students would be equipped with additional sensors attached to the skin. The objective was to use physiological parameters and state-of-the-art approaches as ground truth. Unfortunately we couldn't find students who were willing to wear such devices all day long for a longer time period. So the data recording could not be performed. In contrast, the idea of using only smartphones was widely accepted.

5.7 Behavioral Parameters and the Influence of Medium-Term Stress Periods - Experiment Evaluation

So far behavioral parameters have been defined that are assumed to correlate with medium-term stress periods. This section gives an evaluation based on the data set introduced. Therefore, behavioral parameters were calculated for both time periods and compared to each other. It is clear, that behavior deviations can't be interpreted in the same way for each individual. Each person has specific habits that are represented by behavioral patterns and consequently stress-related deviations can be expressed in different ways. For example, some people are used to having a social life during their free-time and others prefer to live isolated and secluded lives even during stress-free time periods. Hence, the same feature can be very useful for some individual groups for detecting stress-related behavioral changes and can be almost useless for others. In the following, behavioral parameters were considered for each individual and average behavior deviation values were used in order to rank the importance of each feature.

5.7.1 Location Behavior

As was already introduced two location behavior features were defined in this work. The first one (ROI1) includes information about ROIs that were mainly visited during stress-free or stressful periods. Figure 5.3 shows an example for one of the participants.

[61] The application was developed and provided by Jens Weppner, DFKI Kaiserslautern, Germany

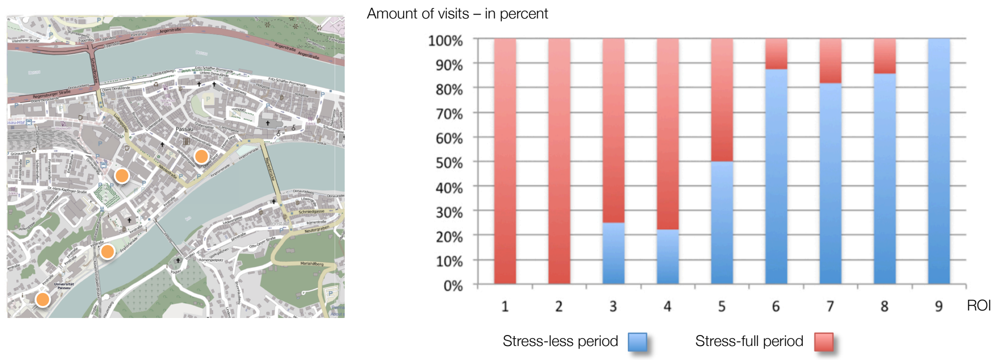


Figure 5.3: Left: ROI findings (orange dots) for one of the participants in the experiment in the city of Passau, Germany (Map created from OpenStreetMap-Data; License: Open Database License (ODbL)). Right: Examples of a participant's ROI visits (in percent) during stress-free and stressful periods.

Four locations (1-4) were almost exclusively visited during the stressful periods whereas another four locations (6-9) were mainly visited during stress-free periods. This shows, that the participant changed his location behavior significantly due to a medium-term stress period. Location 5 was visited in both cases and consequently it is very likely that this ROI represents the person's home. Figure 5.4 shows the amount of important regions (as was defined in section 5.5.1) in percent. Considering all participants, it can be seen that on average 39% of the ROIs

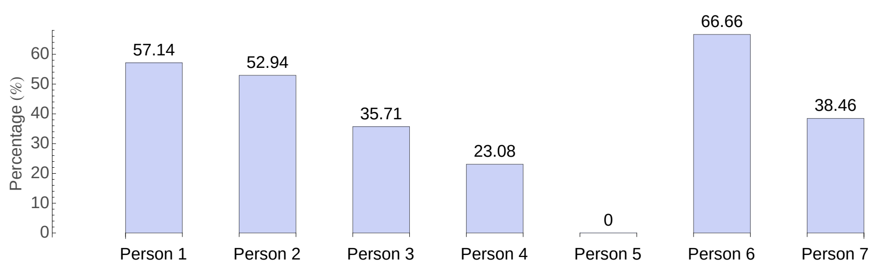


Figure 5.4: Evaluation of ROI1: Amount of important ROIs for each participant in percent (related to found ROIs). ROIs are important according to the definition of feature ROI1 (visited at least twice as often within one period than during the other period and an average retention time of at least 20 minutes).

found were only important (according to the definition of ROI1) during one of the periods (stressful or stress-free). The standard deviation is 21 and the confidence interval is [18; 60] with $(1 - \alpha) = 0.95$. Having a more detailed look, it can be seen that some participants (1, 2 and 3) showed quite a different location behavior (more than 50%) than others (3, 4 and 7). For one of the participants (5), the location behavior was almost the same during both periods according to the definition of ROI1.

Figure 5.5 shows evaluation results in terms of ROI2, which considers the amount of time a person spent at a specific location during both periods. Consequently, the objective was to see whether people spend more or less time at specific locations during a medium-term stress period. On average each participant showed a quite significant behavior deviation of 73%. The standard deviation is 25 having a confidence interval of [48;98] with $(1 - \alpha) = 0.95$. A more detailed look on single participants highlights, that almost all people had a significant behavioral change of more than 70% (which means that the duration of stay differs significantly for the same location within both time periods). Only participant 5 showed almost no deviation (below 15%). Hence, the overall location behavior of this person was not at all influenced by the medium-term stress

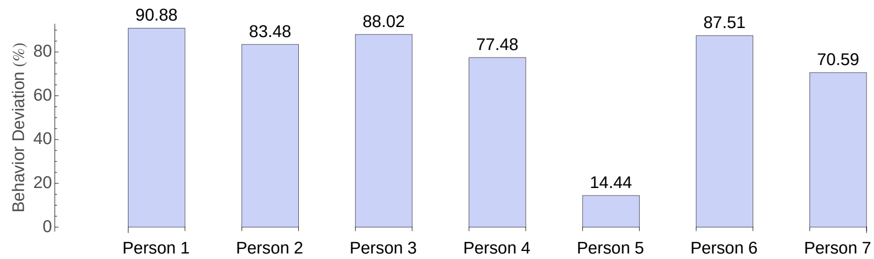


Figure 5.5: Evaluation of ROI2: Average absolute time deviation for participant's ROIs in percent.

period as he already showed no significant deviation for ROI1.

5.7.2 Social Interaction

Two features were introduced to measure the amount of social interaction. Feature SI1 covers the overall amount of social interaction a person has. Figure 5.6 shows how many people were located in the close vicinity of one of the participants during the stress-free and stressful periods grouped by weekdays. It can be clearly seen, that this person was significantly more isolated on weekdays (Wednesday, Thursday, Friday) during the stressful time period in contrast to the stress-free period. The measured absolute deviation from a participant's behavior (related to the definition of SI1) during the stress-free time period is shown in Figure 5.7.

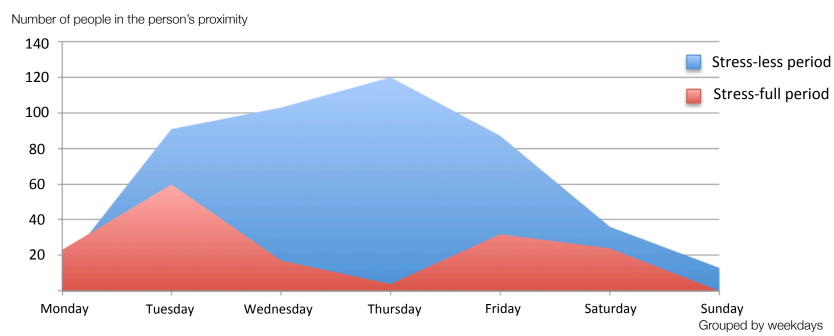


Figure 5.6: Amount of people located in the close vicinity of a participant during the stress-free and stressful period grouped by weekdays.

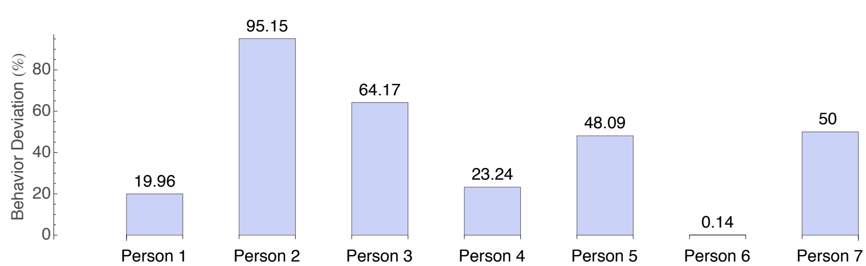


Figure 5.7: Evaluation of SI1: Grade of overall social interaction. Absolute deviation from the stress-free behavior during the stressful period in percent.

On average each participant changed his behavior by 43 percent with a standard deviation of 29. The confidence interval is [14; 72] with $(1-\alpha) = 0.95$. It can be seen, that this feature is quite person-dependent. Some participants (such as 2, 3 and 7) clearly showed different behaviors whereas others (such as 1, 4 or 6) showed almost no behavioral change in terms of feature SI1.

Feature SI2 described the time a person spent with other "important" persons during both periods. The idea was that people would be in contact with different persons (e.g. work colleagues and friends) during stress-free and stressful periods. Figure 5.8 shows the absolute deviation from their normal behavior in terms of SI2 during the stress-free period. Some participants such as 1, 2 and 5 showed a significant behavioral change of more than 78% in terms of feature SI2. Especially participant 2 changed his behavior completely, which means that he spent time with totally different "important" persons during both periods. Other participants like 3, 4, 6 and 7 did not really change their social interaction behavior (below 40%). On average a behavioral change of 55% could be measured for all participants. The corresponding standard deviation is 31 and the confidence interval is [24; 86] with $(1-\alpha) = 0.95$.

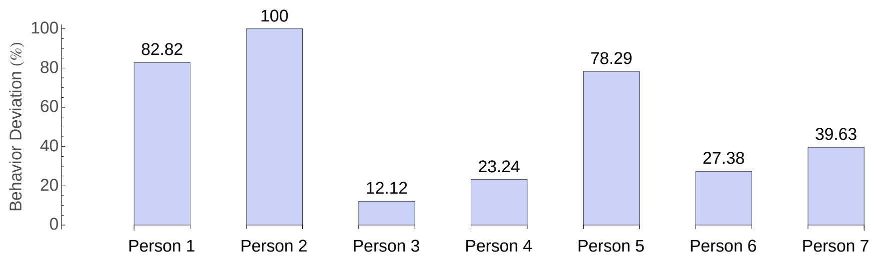


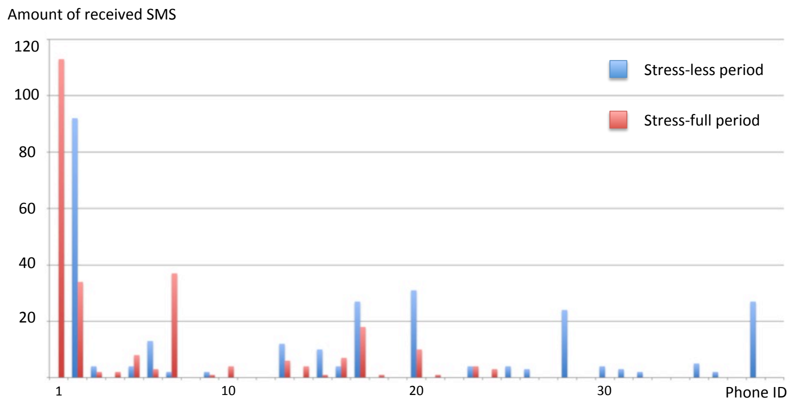
Figure 5.8: Evaluation of SI2: Time spent with important people. The average absolute deviation from the stress-free behavior during a medium-term stress period is shown in percent.

5.7.3 Phone Behavior

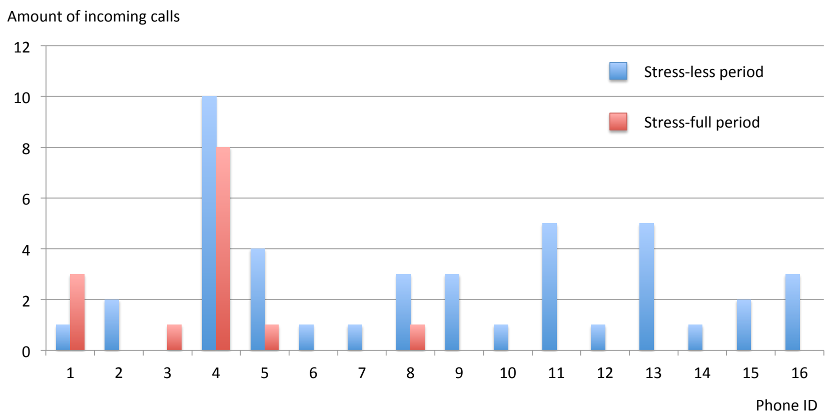
As a last feature group the phone behavior of people during stress-free and stressful time periods was analyzed related to phone calls and SMS. Considered features included the amount of phone calls performed and SMS received as well as which people were contacted. Figure 5.9 shows the amount of SMS received and incoming phone calls for one participant. A clear change can be seen for both cases. Considering the amount of SMS received, it can be seen that phone id 1 and 7 were in strong contact with the participant during the stressful period only. Besides that, several phone id's (e.g. 20, 28 and 38) contacted the participant quite often and mainly during the stress-free periods. In terms of incoming phone calls a similar behavioral change can be seen. Several contacts (e.g. 9, 11, 13, and 16) called the participant several times only during the stress-free periods. This could indicate for example that people knew about the high stress level of the participant and consequently, they tried not to contact him/her in order to make appointments.

Feature C1 and S1 describe the absolute deviation of phone calls performed (incoming as well as outgoing) and SMS received. Figure 5.10 shows the evaluation results. It turns out, that in the case of SMS received only one person (4) showed a significant behavioral change of nearly 70%. On average every participant showed a behavioral change of 24%. The standard deviation is 22 and the confidence interval [33; 86] with $(1-\alpha) = 0.95$. In the case of phone calls four participants showed a behavioral change of 50% and more. On average a behavioral change of 54% with a standard deviation of 21 was detected. The corresponding confidence interval is [33; 75] with $(1-\alpha) = 0.95$.

Besides the amount of phone calls performed or SMS received, this work also analyzed who participants were in contact with. C2 and S2 consider the amount of people that were exclusively contacted during one period (either stress-free or stressful). Figure 5.11 shows the results achieved. It can be seen that all participants except person 2 showed a significant



(a) Amount of received SMS grouped by phone id during both periods for one participant.



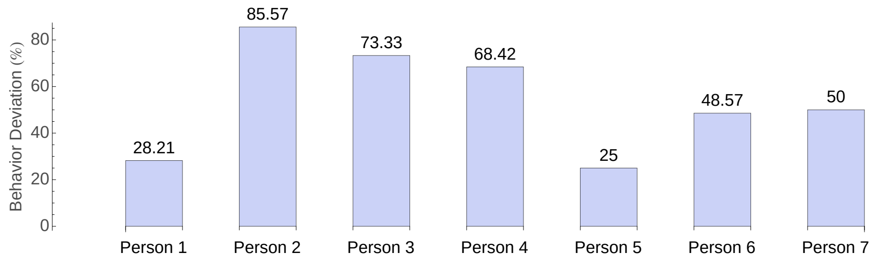
(b) Amount of incoming phone calls grouped by call id during both periods for one participant.

Figure 5.9: Received SMS and incoming phone calls for one of the participants.

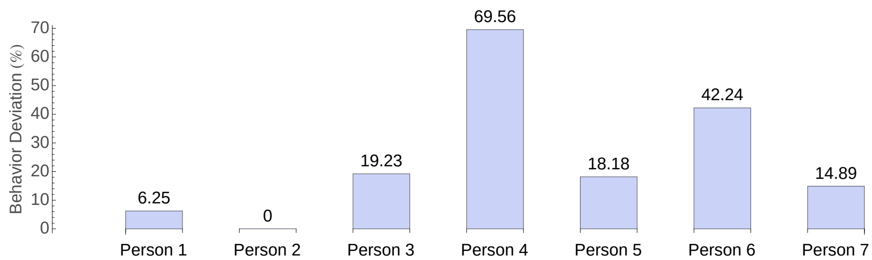
behavioral change of more than 57%. Participant 7 even changed his behavior completely. This means, that for this person completely different contacts were important during both periods. On average a behavioral change of 73% could be seen with a standard deviation of 18. The confidence interval is [54; 91] with $(1 - \alpha) = 0.95$. In the case of SMS almost all participants showed a behavioral change of more than 50%. The average value is 60% having a standard deviation of 27. The confidence interval is [33; 86] with $(1 - \alpha) = 0.95$. It is noteworthy, that person 2 had no contacts that were exclusively contacted during one period.

5.7.4 Summary

This section summarizes the results achieved. In Table 5.1 the behavioral change for each participant-feature pair is shown in percent. It shows, that the behavioral parameters introduced are able to detect a behavioral change of 53% on average. Taking a more detailed look at single participants, one can see, that the average behavioral change of each participant is between nearly 40% and 60% with standard deviations ranging from 22% to almost 38%. The results also show, that features are not generic but user-dependent. SMS-related features for example are very important in order to express behavioral changes for person 4, whereas they are almost useless for person 2. It is obvious, that not every small behavioral change is related to a medium-term stress situation. But even if a minimum average behavioral change of 50% is used as an indicator for periods of stress, the proposed system is able to recognize nearly 86% of our

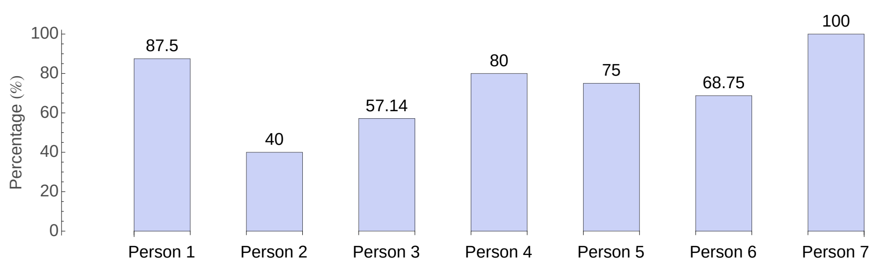


(a) Absolute deviation in terms of phone calls performed during both periods in percent.



(b) Absolute deviation in terms of SMS received during both periods in percent.

Figure 5.10: Evaluation of C1 and S1.



(a) Percentage of call contacts that were important (according to the definition of C2) in only one of both periods.

(b) Percentage of contacts sending SMS that were important (according to the definition of S2) in only one of both periods.

Figure 5.11: Evaluation of C2 and S2.

participants as stressed during the stressful time period. Table 5.2 summarizes again the results achieved per feature. When considering the average deviation for each feature-participant pair, one can see, that feature ROI2 (related to time spent in important ROIs), C2 (related to phone call contact behavior) and S2 (related to people contacting the participant by SMS in only one of the two periods focused on) are very useful to detect behavioral changes in general. On the contrary, feature ROI1 (related to visited locations) and S1 (related to the amount of SMS received) are not able to detect significant behavioral changes on average. However, as already mentioned, each individual is different and as the participant list is quite small, a statement on what type of features are able to detect behavioral changes due to medium-term stress situations in general cannot be given at this point.

Table 5.1: Overall behavioral changes for each participant in percent.

Feature/Participant	1	2	3	4	5	6	7
ROI1	57.14	52.94	35.71	23.08	0	66.66	38.46
ROI2	90.88	83.48	88.02	77.48	14.44	87.51	70.59
SI1	19.96	95.15	64.17	23.24	48.09	0.14	50.00
SI2	82.82	100.00	56.49	12.12	78.29	27.38	39.63
C1	28.21	85.57	73.33	68.42	25.00	48.57	50.00
C2	87.50	40.00	57.14	80.00	75.00	68.75	100
S1	6.25	0	19.23	69.56	18.18	42.24	14.89
S2	75.00	0	85.71	81.92	60.00	56.00	60.00
Average	55.97	57.14	59.98	54.48	39.88	49.66	52.95
StDev	31.30	38.12	22.12	27.64	27.64	25.33	23.53

Table 5.2: Average behavior deviation, standard deviation and corresponding confidence intervals with $(1-\alpha) = 0.95$ for each feature.

Feature	Average Deviation	StDev	Confidence Interval
ROI1	39%	21	[18; 60]
ROI2	73%	25	[48; 98]
SI1	43%	29	[14; 72]
SI2	55%	31	[24; 86]
C1	54%	21	[33; 75]
C2	73%	18	[54; 91]
S1	24%	22	[2; 16]
S2	60%	27	[33; 86]

5.8 Conclusion

This section described the idea of using mainstream smartphones in order to detect behavioral changes of individuals due to medium-term stress periods. Behavioral parameters were introduced in order to describe people's routines in terms of location, social interaction and mobile phone-based communication. It was shown, that these features are powerful enough to reflect behavioral changes which occur during medium-term stress periods. It is clear, that the way in which people change their behavior is specific and different for each individual. Consequently, not all features are useful for all people in the same way. However, it was shown that behavioral features introduced were able to detect a behavioral change of 53% on average. Of course it has to be kept in mind, that the amount of test persons as well as the duration of the experiment performed are not sufficient to reliably prove the results achieved. Nevertheless, first results are very promising and give a good starting point for further and more detailed research work. When summarizing the results achieved, it becomes apparent, that people show different behavior during stress-free and medium-term stressful time periods and that this deviation can be detected by analyzing sensor data from mainstream smartphones. Besides that, the smartphone is still usable as a normal phone in respect to battery life, performance and function. However, in order to achieve more reliable results, ongoing work must be done considering the following points:

- A very important aspect is to increase the number of behavioral parameters. Besides information about outgoing SMS, the length of text messages in general, the duration of phone calls, information about a user's physical activity intensity (e.g. modes of locomotion detection), patterns related to Apps or mobile web usage can be very helpful. Many approaches are already dealing with stress detection based on voice analysis (e.g. [HLMA09] [ZMTA09] [RHM+02]) and thus it is obvious to integrate these algorithms to analyze phone calls.

- It was shown, that single behavioral parameters are able to reflect behavioral changes. As a next step, the combination of several behavior features should be analyzed in order to create a more complex user behavior description.

- As was already mentioned, both the number of participants as well as the experiment's duration are not big enough to prove the results achieved reliably. Hence, the data recording should be repeated by using more participants (10 to 15) and for a much longer time period (several months).

- The assumption that students are stressed during time periods in which they have to pass important but difficult exams is obvious. However, a more reliable and individual specific ground truth must be available to separate stress-free from stressful time periods. One possibility could be to use self-report questionnaires (realized by a smartphone App) as was done in [JPS+13].

Another idea for further research work would be to bring behavioral and physiological parameters together. As physiological patterns are highly influenced by various parameters (e.g. intense physical activities), behavioral parameters can be used to get detailed information about the current scenario. Having information about current user activities like driving a car or being in a very crowded place for several hours, might result in a better understanding of physiological parameters. [GPT+] follows a similar idea, where acceleration sensors are used to get information about the user's activities. Because ECG signal processing techniques are strongly influenced by physical activities, they are only analyzed during low activity time ranges.

Concluding it can be said, that even if the approach introduced is based on initial experiments only, it was shown that defined behavioral parameters are able to reflect user behavioral changes. However, the exact interpretation and generalization requires further and large scale experiments for which the proposed approach is definitely a good starting point.

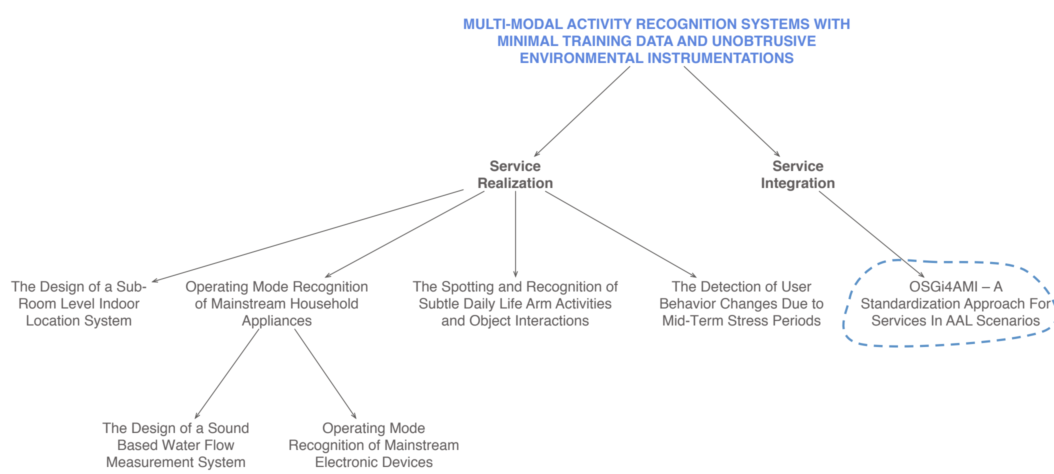
6

OSGi4AMI – A Standardization Approach for Services in AAL Scenarios

This chapter is based on my work published in

Alvaro Marco, Roberto Casas, Gerald Bauer, Ruben Blasco Marin, Angel Asensio, Bruno Jean-Bart and Miriam Ibanez. *Common OSGi Interface for Ambient Assisted Living Scenarios*, in Björn Gottfried and Hamid K. Aghajan, ed., 'BMI Book', Volume 3, Ambient Intelligence and Smart Environments, IOS Press, 2009, pp. 336-357.

Gunnar Fagerberg, Antonio Kung, Reiner Wichert, Mohammad-Reza Tazari, Bruno Jean-Bart, Gerald Bauer, Gottfried Zimmermann, Francesco Furfari, Francesco Potorti, Stefano Chessa, Michael Hellenschmidt, Joe Gorman, Jan Alexandersson, Jürgen Bund, Eduardo Carrasco, Gorka Epelde, Martin Klima, Elena Urdaneta, Gregg C. Vanderheiden and Ingo Zinnikus. *Platforms for AAL Applications*, in Paul Lukowicz; Kai S. Kunze and Gerd Kortuem, ed., 'EuroSSC', Springer, 2010, pp. 177-201.



6.1 Introduction

So far this thesis has focused on the realization of activity and context recognition services for real-life applications. Systems related to user localization, the monitoring of mainstream household devices, daily-life activity spotting as well as behavior analysis have been designed. However, each approach introduced is so far a stand-alone system. Similar to these services, a large variety of pervasive computing systems have been designed and successfully evaluated within living labs and real-world environments over the last few years. Unfortunately, only a few have already found their way into the end-user market. Examples are Nike+ products[62], which monitor users while running, playing basketball or operate as a personal trainer at home. Besides that, RWE smart home products[63] are offered to monitor windows and doors as well as to detect presence in rooms. On-going research will make more complex and intelligent services realizable. Based on the fact that such systems are becoming accepted more and more by society, they will become ubiquitous in the future. As a consequence, the related market is expected to grow significantly.

Due to this fact, it is even more important to overcome existing problems related to the integration of activity recognition services in real homes on a large scale. First of all, adequate residential gateways (RG) have to be found, which are powerful enough to process sensor data and are energy-saving at the same time. Even more important is the selection of a standardized RG framework, which is used to realize and to manage smart services. But clearly, when it comes to service development, extension and integration, one of the most difficult problems is related to the existing heterogeneity among technologies used and services realized.

This chapter provides a feasible solution for this problem by defining standardized interfaces for pervasive computing services in smart home environments. The service-oriented framework OSGi[64] (see [OSG07] [TdOV08]) has already been widely used as a RG platform in many smart environments. Consequently, it was also considered as a RG framework in this work. Besides that, the proposed standardization approach is shown using the example of Ambient Assisted Living (AAL) scenarios. The reason therefore is the fact that this work was mainly driven by the EU project MonAMI[65], which targeted the integration of AAL services in real end-user homes on a large scale.

In the following, related work and the main contribution of this chapter are shown. Afterwards the common structure of smart homes is explained followed by a short discussion about the requirements of residential gateways and OSGi as suitable RG platform. In Section 6.6 common service components in ambient intelligence scenarios are defined. On the basis of these components a standardization approach (called OSGi4AMI) for technical and functional (end-user) services is introduced. Amongst others, activity recognition systems and concepts introduced in Chapter 2 and Chapter 3 were addressed. The statement, that standardized interfaces are a feasible solution for the realization of distributed service developments and the re-usability of existing services is confirmed in Section 6.8. There, the development of two AAL services based on OSGi4AMI and existing components is shown. Finally, these services were deployed in altogether 54 real homes of elderly and disabled people for several month. End-user evaluations showed that they fulfill the requirement of an unobtrusive integration and that they have contributed to a valuable improvement of the end-user's quality of life.

6.2 Related Work

A first approach of standardizing devices was performed by the UPnP forum. However, so far UPnP covers mainly audio and video devices instead of home automation devices in general. Consequently only a few services related to smart home applications have been taken into account. More details can be found on the UPnP webpage[66]. Another standardization approach

[62] http://nikeplus.nike.com/plus/ (last accessed on 2013/05/10).
[63] http://www.rwe-smarthome.de/ (last accessed on 2013/05/10).
[64] http://www.osgi.org/ (last accessed on 2013/05/10).
[65] http://www.monami.info/ (last accessed on 2013/05/10).
[66] http://www.upnp.org/standardizeddcps/default.asp (last accessed on 2013/05/13).

was done by Telefonica I+D (see [Iba02] [MVMJMC03]). There, a three-level architecture was proposed consisting of a low-level network layer (the device is an element within the network), a device layer (devices are grouped regarding their functionality) and an application layer (all devices share the same interface regardless of their type). A similar concept of a three-level architecture was also introduced in [JRAJL04]. In [LNZ12] a five layer architecture is shown. There, a low level sensor layer distinguishes between wearable and static sensors. An additional transmission layer is used to import data from different sensor networks into the gateway. Afterwards raw sensor data is processed using data analysis and pattern mining layers. Based on processed data, services are implemented in the following application layer. Finally, a security layer handles privacy issues such as data encryption and authorized access for each layer. The definition of the ZigBee Cluster Library (ZCL) has a similar approach to group functionalities of devices in clusters and attributes (see ZigBee Cluster Library Specification[67]) as is done in OSGi4AMI. There, clusters are grouped into domains addressing specific functionalities (e.g. security, lighting, etc). Although OSGi4AMI also introduces clusters and domains, ZCL focuses on the driver level whereas OSGi4AMI defines high-level layers. In [HWL+03] a first approach was made to create a re-usable ultrasonic based user tracking service. The objective is to provide third-party developers an easy way to integrate the location service in their own applications. Therefore a single OSGi bundle was created that hides all related system components ranging from sensor communication functionalities over data processing to the derived user location system. However, a standardization approach was not proposed and hence the re-usability of single components is not given. One method to harmonize data from different sensor networks is shown in [FMH+10]. There, a hardware abstraction layer is used to define a common data format. Non-standard devices are mapped to appropriate notations in the ISO 11073 specification. However, a detailed description of AAL services is not considered and data analysis is performed by a common event processing bundle. In [VF02] additional components for OSGi were introduced including basic bundles used to realize applications, a home portal extension as well as a remote management system. Basic components handle topics like device access (support for various hardware protocols), internet access (like POP3) or security support (e.g. SSL). The home portal extension is mainly based on a macro engine. This component was designed for non-technical users and provides a simple method to program operation sequences. An approach to create high-level services based on existing OSGi bundles is introduced in [RVC+07]. There, the issue of OSGi service composition based on atomic services at run-time is handled. The composition is done by a transparent BPEL (orchestration language for web services) style solution. Although the work deals with the re-usability of existing bundles, services of the same type and services providing similar information can still be designed in a completely different way. The problem of heterogeneity of standards and protocols in home networks is covered by [CT09]. Besides that, authors focus on the resulting problem of an easy and automated device discovery, registry and management. Finally many approaches deal with OSGi extensions for various application areas. Examples are [HLS+12] and [HC10]. There, extensions for OSGi are shown in order to integrate it into cloud computing applications and to provide instant messaging communications and peer to peer transfer. In [WWS07] and [WDJ10] OSGi add-ons for RFID applications are shown. However such extensions are not related to the work shown in this thesis, they illustrate again the importance of OSGi as a service platform and its relevance for smart home applications.

6.3 Research Questions and Contribution

This chapter faces the problem of missing standards for pervasive computing systems in smart home environments. Due to the lack of standardization, service providers are completely free to realize smart services and hence interfaces for common devices, low-level as well as high-level services differ from service provider to service provider. Apart from that, commercial systems use closed and proprietary protocols to a large part. As a consequence, existing services and systems can hardly be reused and integrated in new smart home environments. This fact makes

[67] http://zigbee.org/Home/SearchResults.aspx?q=cluster+specification (last accessed 2013/05/14).

a distributed and efficient development of new systems hard or even impossible.

Consequently, the ***key problem*** can be described by:

How can ambient intelligence services in smart home environments be standardized in order to make a distributed service development as well as a straightforward re-use of available services possible?

This chapter starts with a basic discussion about smart homes, residential gateways and OSGi, which has been widely used as gateway framework in such scenarios. Hence, the ***first research question*** lays the foundation for the proposed standardization process.

Is the widely used OSGi framework really the best solution for residential gateways in smart home environments?

In Section 6.5 a treatise on OSGi and requirements on residential gateway frameworks is presented, which confirms its usage in smart home scenarios. Based on OSGi, the ***second research question*** is related to the structure of OSGi and smart home services in general.

What are common service components in ambient intelligence scenarios and how can they be integrated into OSGi?

In Section 6.6 a three layer architecture is proposed. There, ambient intelligence services were grouped into low-level devices/sensors, basic technological services and high-level, end-user functional services. Due to the integration into OSGi, functionalities provided that are related to service deployment, management and communication were automatically re-used. Based on defined components, the ***third research question*** is related to the main objective of this chapter:

How can ambient intelligence services for smart home scenarios be standardized in OSGi by considering aspects of service re-usability, distributed deployment and appropriate scope for creativity related to service realization?

In Section 6.7 a feasible standardization approach is introduced. In order to guarantee enough creative freedom in the case of service realization, only technological services are standardized. This service category includes data processing as well as recognition services and provides high-level context information to other services. As a next step, technological services were grouped into application categories. Based on these considerations, standardized interfaces were defined. The key idea was, to use key-value lists in order to enable both: Pre-defined access on standard methods and features as well as providing enough freedom to extend services and to adapt them to specific scenarios.

The resulting set of interface descriptions (called OSGi4AMI) was provided as an extension for OSGi. It includes interface descriptions for more than 30 common technological services. The ***fourth*** and last ***research question*** is related to OSGi4AMI and its use in real-life scenarios.

Is OSGi4AMI a suitable solution to realize and deploy pervasive computing services related to ambient assisted living services in real homes of end-users on a large scale?

In Section 6.8 two end-user services related to home and user security were realized based on OSGi4AMI and its advantages of service re-usability. Altogether 54 service instances were deployed in real homes of disabled and elderly people in three countries. Trials have shown, that the proposed standardization approach is a suitable solution to deploy smart services and smart homes in general in real end-user homes. Besides that, surveys (conducted by LSE, London, United Kingdom) proved, that users rely on smart services, they had influenced their ability to view the future optimistically and systems based on concepts, as shown in this thesis, can be unobtrusively integrated into exiting environments.

In summary, the ***key contributions*** of this chapter were:

- The definition and realization of standardized interfaces for technological services in ambient assisted living scenarios as an OSGi extension.
- The definition of common service components in ambient intelligence scenarios, a three-layer architecture (device level, technological and functional services) and their integration into OSGi. (This work was done in co-operation with Alvaro Marco and Roberto Casas (Tecnodiscap Group, University of Zaragoza, Spain))
- The development of AAL services and their large-scale, real-life deployment based on OSGi4AMI.

I would like to note, that OSGi4AMI was evaluated within the EU project MonAMI[68] and was also considered to be integrated in the AAL Open Association (AALOA)[69]. Besides that, the standardization approach was supported by the EU project EasyLine+ (IST-045515) and by the Spanish Ministry of Science and Technology under the AmbienNET project (TIN-2006-15617-C03-02).

6.4 Smart Home Architecture

Typical smart home environments include a variety of smart services, which are based on different types of sensors and actuators. These components send raw or pre-processed sensor data to a central computation unit, where all information is fused in order to realize smart services. Depending on the components used, several different networks using various transmission media (e.g. PLC or wireless) and protocols (e.g. ZigBee, ZWave, LonWorks, KNX and more) can co-exist in a common smart home environment. Sensors and actuators mainly use ZigBee networks to send raw data, whereas smart home appliances are mostly based on PLC networks. Home information like lighting, temperature and heating is usually delivered by bus systems like KNX. Consequently, smart homes can contain several different network technologies. Due to this fact, the central computation unit must be able to handle a variety of sensor networks, it has to be powerful enough to realize smart services (data processing and fusion) and it should be energy-saving at the same time. So called "Residential Gateways" (RG) seem to fulfill these requirements. Residential gateways can be seen as an evolution from current DSL modems or set-top boxes aiming at connecting homes and enabling services[70]. In terms of AAL, the RG is a platform on which assistive services are deployed and offered to end-users. Figure 6.1 visualizes the architecture and common network types within a smart home environment. To sum up, it can be said, that smart homes exist on a variety of different sensor and actuator networks which are connected to a central residential gateway on which smart services are operated and managed.

6.5 Residential Gateways and OSGi

Residential gateways must be able to run software bundles (services) that fuse, analyze and process data from connected home automation sensors as well as to interact with the environment, actuators and end-users. As it is one of the main objectives to have a common platform which is able to offer a portfolio of services coming from different service providers, an open and common execution environment must be defined. In many application domains like embedded systems or mobile phones, the Java based service platform OSGi (Open Service Gateway Initiative) (see [OSG07] [GVG07]) is used as a RG platform. So far many research projects related to smart

[68] http://www.monami.info/ (last accessed on 2013/05/10).
[69] http://www.aaloa.org/ (last accessed on 2013/05/10).
[70] http://www.homegatewayinitiative.org/ (last accessed on 2013/05/11).

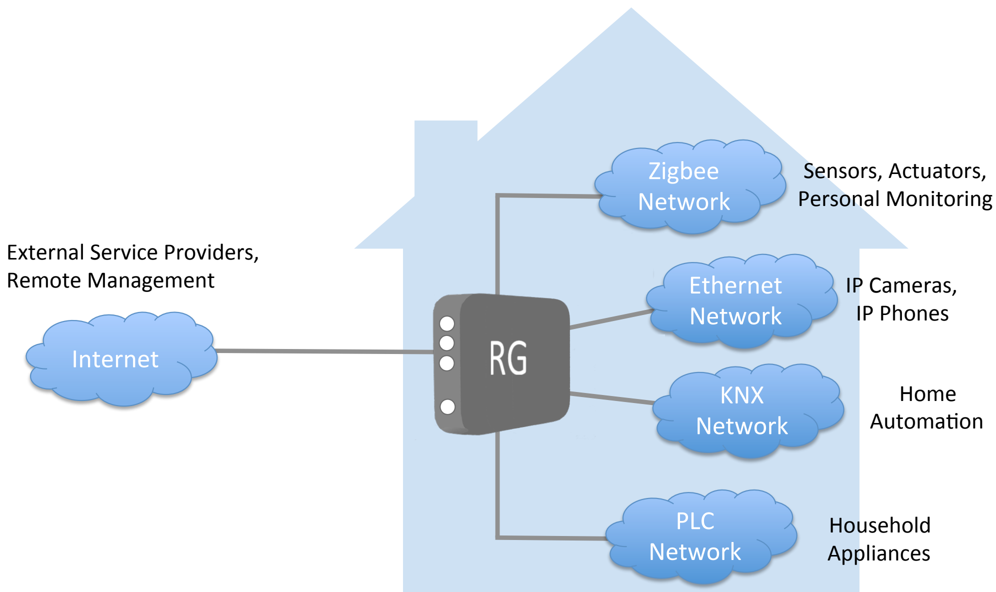


Figure 6.1: Smart home architecture: Smart homes may consist of several sensor networks like ZigBee, KNX or PLC. The residential gateway (RG) is a central processing unit where data packets are processed and fused. Consequently, the RG is a platform on which smart services are deployed and offered.

environments have used OSGi as a RG platform. Examples are MonAMI (www.monami.info, 2009), EasyLine+ (www.easylineplus.com, 2009), Aspire (http://fp7-aspire.eu, 2009), ePerSpace (http://www.ist-eperspace.org, 2009), Amigo (http://www.amigo-project.org, 2008), AmiVital (http://www.amivital.es, 2007), TEAHA (http://teaha.org, 2008). Share-it (http://www.ist-shareit.eu/shareit, 2009) or H@H (http://www.hearing-at-home.eu/, 2009). Besides, as it was already mentioned in Section 6.2 much research work was done on extending and adapting OSGi to specific application scenarios. OSGi is also used in many different systems like the development environment Eclipse[71] or in embedded systems such as the BMW 3 series car. These facts confirm that OSGi is a widely used solution for residential gateway platforms today. As the proposed standardization approach was realized as an extension for OSGi, the concept and idea behind this framework is briefly introduced in the following. Besides, it is discussed why OSGi was chosen as a service platform and why it is a suitable solution for residential gateways.

6.5.1 OSGi Framework

OSGi is based on the SOA (Software Oriented Architecture) model. It provides a framework used to deploy so called OSGi bundles, which represent services. An OSGi bundle is in fact a software component based on Java classes providing functionalities (e.g. realization of services) and including OSGi-related information. The framework provides methods to manage the lifecycle of bundles as well as the interoperability between specified bundles. Consequently, bundles can share information among other installed and registered bundles. Imagine a service bundle that was designed to retrieve information from motion sensors. Such a bundle has of course to implement the corresponding motion sensor protocol. Another bundle can now register to the motion sensor bundle and in this way it will be informed about detected motions without any knowledge about the underlying sensor system and the required protocol. Another big

[71] http://www.eclipse.org (last accessed on 2013/05/11).

advantage of OSGi is the ability to start, stop or to update bundles online without restarting the whole system or unaffected bundles. Besides that, OSGi can even be controlled remotely. In summary, OSGi provides a powerful framework for real-time management of service bundles. More detailed information about OSGi functionalities can be found in [OSG07] [TdOV08]. In the following requirements of a RG service platform and the usability of OSGi in smart home scenarios are discussed.

6.5.2 Requirements on a RG Service Platform

As was already mentioned, the main objective of the RG is to connect internal networks, realize smart services and handle communication with the external world. Consequently, RG platforms have to provide an open development environment to deploy services. Smart home environments are designed to be very dynamic. This implies, that new services and related sensor technologies should be capable of being integrated into the existing environment with ease. Hence one of the most important requirements on RG service platforms is the ability to handle new services, sensors and actuators as well as new network technologies in real time and remotely. The vision is that people can rent a basic system including the RG and that they are able to book assistive services online. Hence, new services must be installed and integrated into the existing system remotely. In addition, new sensor systems and related network technologies must be easily installable or preferably working out-of-the-box. These aspects are very important to guarantee real-life scenarios on a large scale. Another important issue is, that the chosen service platform should be independent on used architectures as well as operating systems. Furthermore, the platform should be able to run on heterogeneous computation units and especially on those with slow processors and little memory. In large-scale, real-life applications, residential gateways must be low-energy and low-cost in order to survive on the market. Based on these facts, the following section discusses if OSGi is a suitable service platform solution for smart home environments.

6.5.3 OSGi, a Suitable Solution for RGs?

SOA platforms, which are widely implemented based on web services, fulfill many of the mentioned requirements in terms of easy service development and dynamic systems. Although web services are able to operate between different machines and platforms, the corresponding message exchange mechanism involves a not insignificant time delay. For many applications such as bank transactions a delay of a few seconds is still acceptable. However, in critical smart home applications it is not. When considering that many assistive services are based on several chains of data processing and message exchange bundles, the importance of a fast and immediately service communication becomes even more clear. OSGi is based on Java and the SOA model. Consequently, typical SOA functionalities are picked up and included in OSGi and hence it is an open standard for modular Java application development and management. The further development of OSGi is driven by the OSGi Alliance with contributions from IBM, ORACLE as well as device manufacturers such as NOKIA, BOSCH and Siemens (see [RDG+08]). Many publications (e.g. [OSG07] [RVC+08]) mentioned statements featuring OSGi as the most widely adopted technology for building control systems within networked homes. When comparing formerly introduced requirements on a residential gateway with OSGi functionalities, this fact is not astonishing.

OSGi is a Java based platform and consequently, it is platform independent. Besides that, OSGi is a modular approach which offers the possibility of creating basic and stand-alone OSGi components called OSGi bundles. Such bundles realize services ranging from simple sensor drivers to complex assistive services. As bundles can be connected to interoperate between each other, they are able to exchange information almost in real time. Hence, implementation details can be hidden or even changed easily. Besides that, new services can be created based on already existing bundles. The connection between bundles is done by the framework's service registry. A dynamic bundle management system allows to start, stop, install, remove and update bundles during runtime without influencing the whole system. This feature is very important for smart home environments as such systems show a very dynamic behavior. Besides, OSGi can be controlled remotely with ease. All these characteristics feature OSGi to be a suitable solution for RG platforms. However, there are still open issues and disadvantages when using OSGi in smart environment scenarios. One of the most negative issues is the lack of pre-defined interfaces for ambient intelligence services and devices of the same type. Consequently, service and device development among different service providers and hardware manufacturers is very hard and sometimes nearly impossible. The following example should illustrate again the importance of common interfaces. Imagine two hardware manufacturers that are both producing temperature sensors. One provider offers a ZigBee based solution and a simple service which is able to deliver the current temperature by calling the method "`public int getTemperatur()`". The other manufacturer offers Bluetooth based temperature sensors including a service bundle using the public access method "`public double readTemp()`". So far OSGi provides an excellent mechanism to hide specific implementations as the fact that high-level services can simply register to available bundles and ask for the current temperature value. However, high level services have to be adapted to underlying services as there are no standardized ambient intelligence interfaces in OSGi and consequently similar service types are not forced to share the same access methods. Another big issue is the lack of an easy service discovery mechanism. As soon as OSGi bundles want to receive information from other bundles, they have to register with them and consequently software engineers have to specify the exact service names. Of course this is a big restriction in terms of a dynamic service landscape and the re-usability of existing services. In [RVC+08] a semantic solution is proposed to handle this issue. Besides, OSGi has to also face other smart home related problems. However, the research community has already proposed several workarounds and extensions in order to guarantee a perfect usage of OSGi in smart environments. In summary it can be seen, that even if the basic version of OSGi does not cover all problems related to smart home applications, it fulfills the basic requirements on a residential gateway platform. Moreover, when considering the large amount of existing extensions, OSGi is for sure a suitable solution for smart home scenarios focusing on large scale and real-life applications. When comparing OSGi to other frameworks such as Jini, Debian-Package, XBone, SNMP and OCAP, it seems to be the best solution for smart home environments (see [BGL07] [LNH03] [VF02]).

6.6 Common OSGi Components for Service Realizations in Ambient Intelligence Scenarios

In general a residential gateway is connected to various sensor and actuator networks. Consequently, the RG platform needs OSGi components that are able to communicate with available sensors and actuators. Besides that, components must be available which are used to process and to analyze data received and to realize assistive services. In this work a three-level architecture similar to [Iba02] [JRAJL04] is proposed. The architecture definition is based on co-operation work with Alvaro Marco and Roberto Cases (Tecnodiscap Group, University of Zaragoza, Spain).

The architecture consists of a low-level driver layer, a medium-level layer providing basic services (from now on called technological services) and a high-level layer including complex services (hereafter called functional services). Figure 6.2[72] visualizes the proposed three layer architecture. Next each component is explained more specifically.

6.6.1 Drivers and Devices

Driver bundles are located in the lowest layer of the proposed architecture. They implement network protocols and are used to communicate with available devices. In general a single driver bundle must be installed for each network technology. Figure 6.2 shows an example using only two different network technologies. Hence two driver bundles are available of which one is able to handle ZigBee and the other LonWorks based devices. Besides drivers, even devices

[72] The figure was created by Alvaro Marco (Tecnodiscap Group, University of Zaragoza, Spain) and was taken from [MCB+09].

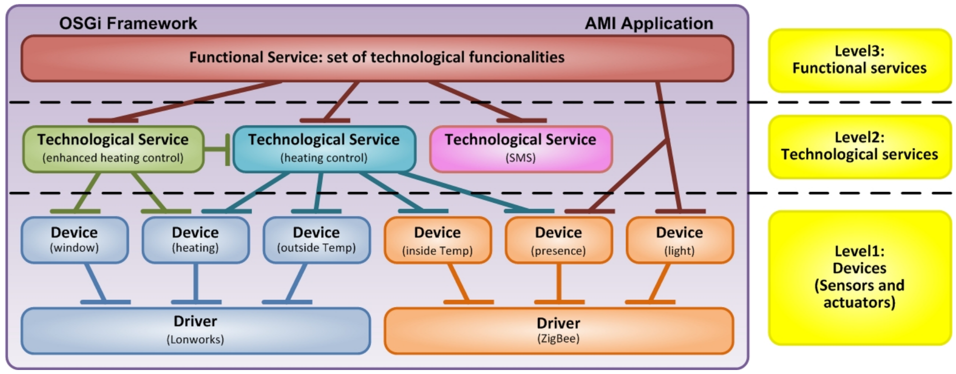


Figure 6.2: OSGi Components: A three-layer architecture (Source: [MCB+09])

are located in the lowest layer. A device bundle is the virtual representation of a physical device within the RG. Such bundles encapsulate device operations and are used to handle two-way communications with the real device (data could be sent to the RG and devices could be controlled by the RG). As data coming from devices hold basic information for services in higher levels, device bundles must provide methods to share their information. Figure 6.2 shows six devices based on LonWorks (e.g. windows and heating) and based on ZigBee networks (e.g. presence sensor and light sensor).

6.6.2 Technological Services

The second layer includes so called technological services. Such OSGi bundles represent services on quite a technical level. Consequently, technological services are used to process raw data coming from devices and to provide basic services (e.g. modes of locomotion detection or indoor localization). Besides that, they are used to control devices (e.g. turn off all lights) or to communicate with the environment (e.g. send SMS). An example is a temperature monitoring service, which processes information coming from temperature sensors and raises an alarm in the case of a too high or too low room temperature (based on pre-defined thresholds). As shown in Figure 6.2 technological services can receive information from several devices at the same time. Moreover, technological services are also able to exchange information among themselves or to inform functional services which are located in the top layer.

6.6.3 Functional Services

Functional services are located on the top layer of the proposed architecture. Such bundles realize high-level and end-user services, which consist of a combination of device bundles and technological bundles. Consequently, the implementation is often based on IF-THEN-ELSE rules. The conditional and operational part of such a rule could combine information from both devices as well as technological services. The following example should illustrate again the concept of a functional service. The objective is to create a smart heating control service. The service should be able to automatically regulate the room temperature. As it is a waste of energy to heat rooms if nobody is at home, the service should regulate the temperature only if people are present. The realization of such a service could be done as follows. First, temperature device bundles and a related driver bundle which is able to communicate with real temperature sensors has to be installed. In the same way, heating control devices as well as presence sensors and related drivers (in the case of different network technologies) have to be installed. Besides that, a technical service ($PersonAtHome$?) must be realized, which registers to presence sensors and is able to detect whether people are currently at home. The new functional service has to register to the technical service $PersonAtHome$?, to temperature devices as well as the heating device. Similar rules, as shown below, can be used to realize the focused $SmartHeatingControl$ service.

$$\textbf{IF} \quad \begin{array}{l} ((Temp_{inside} < Temp_{min}) \text{ OR} \\ ((Temp_{inside} - Temp_{outside}) > \Delta Temp_{max})) \text{ AND} \\ (PersonAtHome? == TRUE) \end{array} \quad \textbf{THEN} \quad \textit{turnOn Heating}$$

$$\textbf{IF} \quad (Temp_{inside} > Temp_{max}) \quad \textbf{THEN} \quad \textit{turnOff Heating}$$

In this case the room temperature should always be above a certain value if people are at home. Apart from that, the heating is also turned on, if the temperature measured outside is much lower than the temperature measured inside. Consequently, a quick room temperature decrease is prevented. Regardless of whether people are at home, the heating should be turned off if the room temperature exceeds a maximum value. Figure 6.3 visualizes again the proposed automatic heating control service. I would like to note, that the realization of such a service can also be done in a different way. For example, temperature comparisons can be moved to a stand-alone technical service. Consequently, the event of a too low room temperature can be re-used by several technical and functional services.

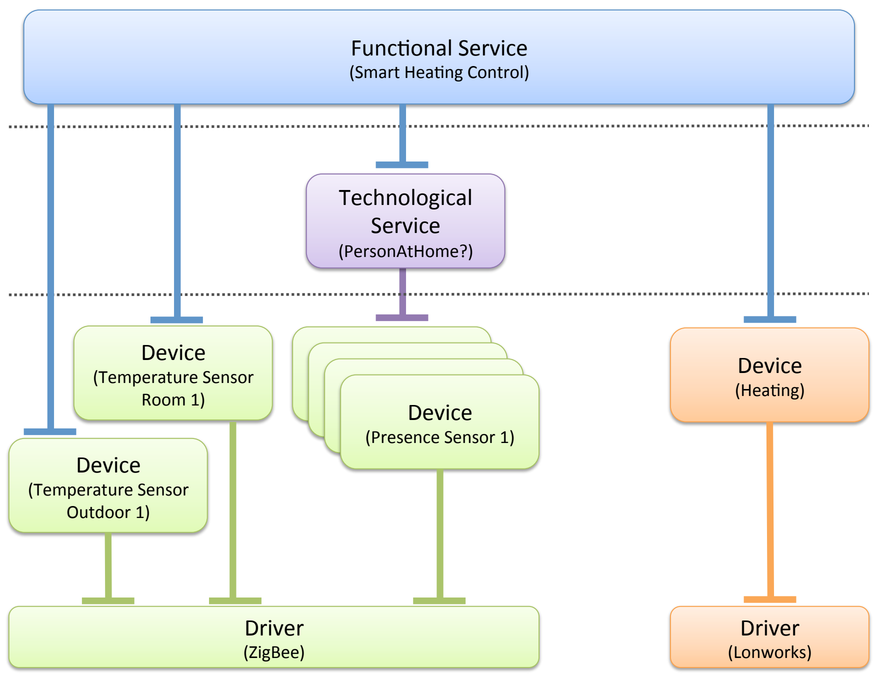


Figure 6.3: Realization of the functional service $SmartHeatingControl$.

6.7 OSGi4AMI: A Standardization Approach for Ambient Intelligence Services and Sensors in OSGi

After the architecture of ambient intelligence applications in OSGi was introduced, this section proposes a standardization approach for components shown. When it comes to standardization, it is very important to guarantee enough tolerance to service developers. Otherwise software engineers become too restricted in terms of creativity. As device drivers depend pretty much on the used hardware and functional services are used to realize complex and high-level end-user services, it would be too restrictive in terms of service development to define interfaces for these components. Thus, standardized interfaces were defined for devices as well as technological services as these components include fairly common elements which occur in many AAL applications.

The standardization process can be done by following several different approaches considering aspects like technology, networking, etc. This work uses the intrinsic nature of devices and services to define interfaces. This means, that common attributes and methods were defined which are valid for all devices or services of the same type. Although resulting standardized interfaces (called OSGi4AMI) should be related to ambient intelligence applications in general, it should be mentioned again, that devices and technological services are adapted to AAL scenarios due to the fact that this work was strongly driven by AAL related projects such as MonAMI (www.monami.info, 2009) or EasyLine+ (www.easylineplus.com, 2009). In [MCB+09], interfaces for devices and technological services were introduced. As the device standardization approach is mainly based on the work of Alvaro Marco and Roberto Casas (Tecnodiscap Group, University of Zaragoza, Spain), it is not covered by this thesis. In the following I will introduce my work which is a standardization approach for technological services.

6.7.1 Technological Services and Application Areas

As a first step, technological services are grouped into bundles in terms of their application area. The following categories which are present in many AAL scenarios are defined.

- **Communication:** This bundle includes services which provide basic communication functionalities between the smart home and the "outside world". Examples are services which can be used to send SMS and Emails or to start phone calls.

- **Personal Monitoring:** This group includes services that provide information about specific users. Such services are of high interest – especially for health care and behavioral monitoring applications. They cover topics such as person localization, mode of locomotion recognition and abnormal behavior or fall detection.

- **Ambient Monitoring:** Besides the monitoring of users, the surveillance of the user's environment is an important issue in AAL applications. Services such as area or region of interest surveillance and the monitoring of common household items are handled in this group. Examples are systems which are able to monitor the operating mode of common home appliances and items such as washbasins, multi-media devices, doors, windows or light and shutter switches.

- **Ambient Control:** Many AAL applications use technologies to control the environment of a user. Assistive applications especially are based on such services. Thus, services able to control or to change the mode of objects placed in the environment are covered by this group. Examples are lights, shutters, doors and windows which can be controlled remotely as well as controllable electronic devices such as smart washing machines or intelligent ovens.

- **Personal Support:** AAL applications are often designed for people with special needs. The aim of such systems is to support people in their daily-life tasks. Hence this group covers services supporting people in terms of their disabilities. Examples are services providing functionalities like speech recognition or text to speech conversions.

- **Special Services:** This category includes services that provide generic support to other services in the framework. Examples are common information about registered users, service usage statistics or special services that do not fit in any of the groups introduced.

Figure 6.4[73] shows service groups proposed and examples for service interfaces realized.

[73] The figure was created by Alvaro Marco (Tecnodiscap Group, University of Zaragoza, Spain) and was taken from [MCB+09]

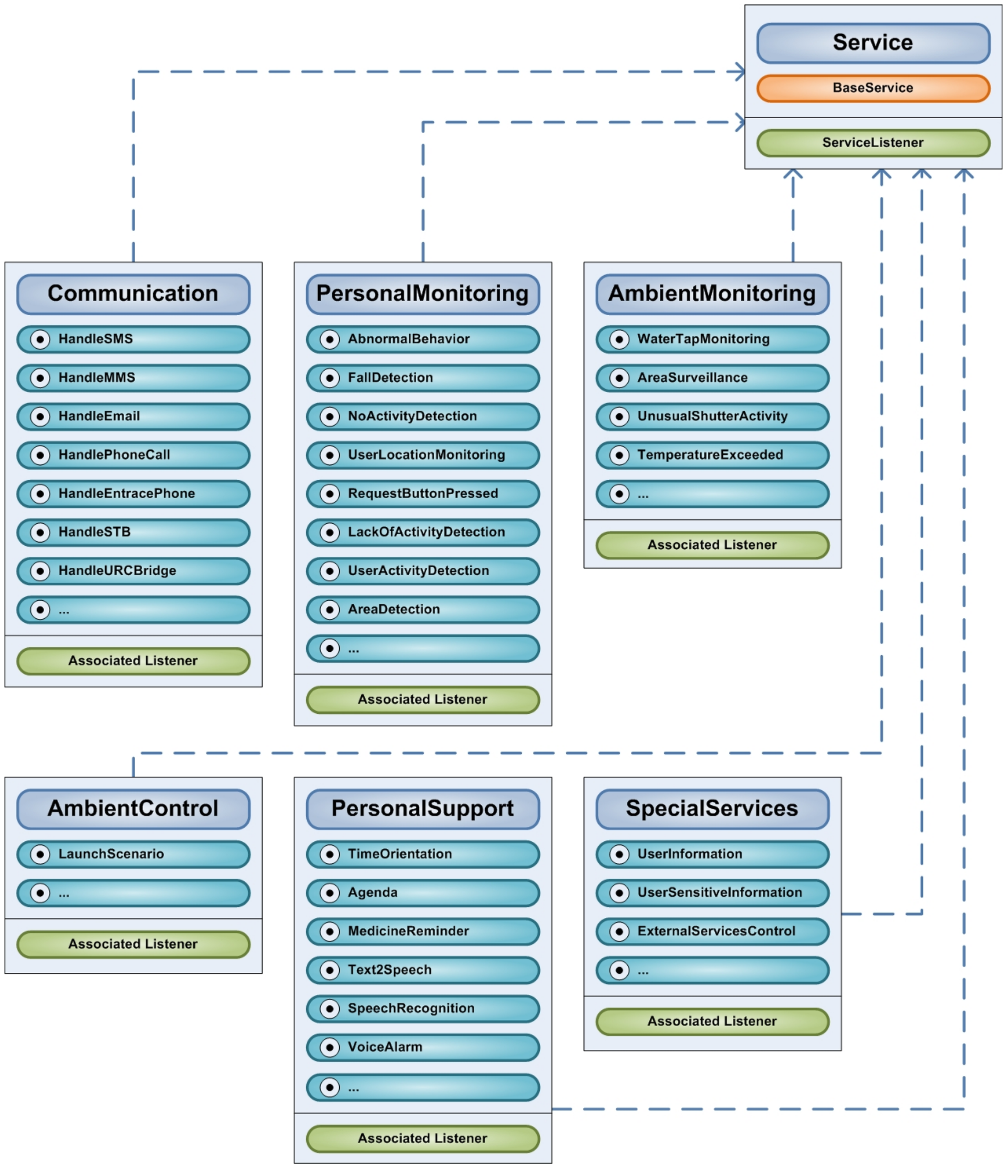


Figure 6.4: Proposed technical service groups and realized interfaces (Source: [MCB+09]).

6.7.2 Standardized Interfaces for Technological Services

This section discusses the realization of the interface standardization in detail. In general, each service consists of two components: The service type interface itself as well as a service specific listener interface. Newly developed services that belong to a specific service type have to implement the corresponding service interface. In this way a common service usage among different service developers can be guaranteed. Services, that want to receive information from another service, have to register to that service and they have to implement the corresponding service listener interface. So, a common method to transmit information among services can be guaranteed.

Before specific technological services are focused on, a generic service interface, is introduced. All technological services inherit from this generic service class and consequently they have a basic feature set in common. Figure 6.5 shows the class diagram of the generic service class. Thus, technological service inherits the following features and methods:

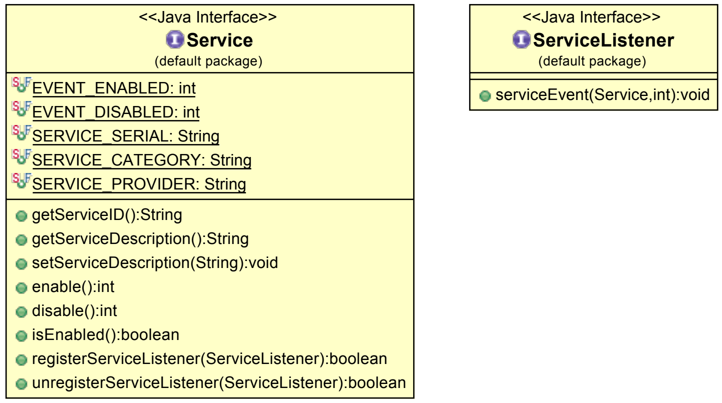


Figure 6.5: Generic service interface: Class diagram.

- **Events:** As smart homes are quite dynamic environments, every service must provide functionalities to be enabled and disabled. This is covered by methods as `enable():int` and `disable():int`. Every time a service is enabled or disabled registered services must be informed that the corresponding service is no longer available. This is done by pre-defined events `EVENT_ENABLED` and `EVENT_DISABLED` which are provided by the service listener method `serviceEvent(Service,int)`. Additionally, services can ask for the current state of a service using `isEnabled():boolean`.

- **General Service Information:** Common service information as `SERVICE_SERIAL`, `SERVICE_CATEGORY`, `SERVICE_PROVIDER` and methods to get and set service descriptions are mandatory for each technological service.

As already mentioned, a common procedure to inform about service events is to implement the corresponding listener interface. Although the way in which services are informed in the case of an event can be realized by standardized methods, it would be too restrictive to standardize the information grade provided. The following example should illustrate this issue. Quite a common problem in AAL applications is user localization. Depending on the underlying technology a more or less accurate location accuracy can be provided (e.g. ranging from room-level location over sub-room level till quite accurate user coordinates). As it would be too restrictive for service developers to assign the reached accuracy of their location systems to pre-defined location grades and consequently accuracy information gets lost, it is not reasonable to provide fixed pre-defined location levels. Of course this issue is portable to all service events and service information. To overcome this problem, this work proposes to code event information in a map container. Consequently, each service can provide standardized key-value entries for common features. In the case of unusual and very specific information, service providers are still able to integrate them by defining their own key-value pairs. Of course, if this solution will dominate the way information is spread between services, the idea of standardization will be destroyed. Nevertheless, this solution provides standardized services with all their advantages in terms of distributed development and service re-usability while still keeping enough freedom to design innovative and specific services.

All in all, interfaces for more than 30 technological services which are widely used in typical AAL applications were designed. As it is beyond the scope of this work to introduce all standardized interfaces and OSGi4AMI is available free on Sourceforge[74], four interfaces are described in the following more specifically to illustrate the idea behind the standardization process.

[74] http://sourceforge.net/projects/osgi4ami/ (2009).

6.7.2.1 Personal Monitoring – User Location Monitoring

First, a user location monitoring service is considered. The objective of this service is to detect if a specific person enters or leaves pre-defined areas. One possible solution for such a sub-room level location service was already introduced in Chapter 2. In general, user location monitoring services have to provide information about what regions were entered/left and who has entered/left them. Therefore interfaces shown in Figure 6.6 were defined.

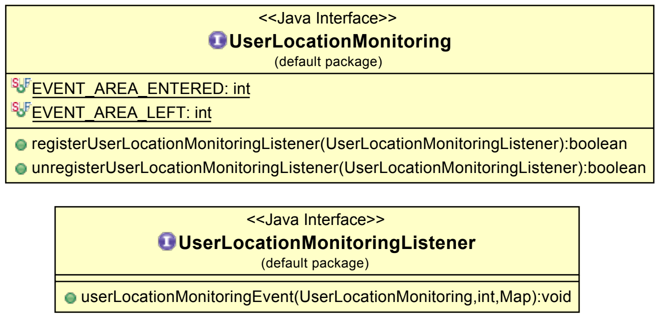


Figure 6.6: User location monitoring service: Class diagram.

Besides methods to register and unregister listener services, the following two events are considered:

- `EVENT_AREA_ENTERED:int`
- `EVENT_AREA_LEFT:int`

Service listeners are informed about events by using `userLocationMonitoringEvent`. Besides the corresponding location service and the event type, a map including optional and more detailed information about the event is provided. Table 6.1 shows pre-defined map key-value pairs.

Table 6.1: User Location Monitoring: Pre-defined key-value pairs.

Key	Information	Format
time	event timestamp	String (yy-mm-dd_hh:mm:ss)
user_id	user identification	String
area_id	area identification	String
area_name	area name	String
area_desc	area description	String
area_2dim_def	area definition	String $(x_1, y_1; x_2, y_2; ...; x_n, y_n)$

6.7.2.2 Ambient Monitoring – Water Tap Monitoring

The second service is related to ambient monitoring. Services as introduced in Chapter 3 provide valuable information about devices located in the user's environment. Based on such services, applications can be realized, which are able to recognize user behavior patterns and unusual deviations. Besides that, assistive services can be deployed which can prevent dangerous situations such as left on water taps. The latter mentioned issue is considered here. Figure 6.7 shows standardized interfaces for a water tap monitoring service.

A common `WaterTapMonitoring` service should be able to provide information about the current status of the water tap as well as information about the current water flow in the case

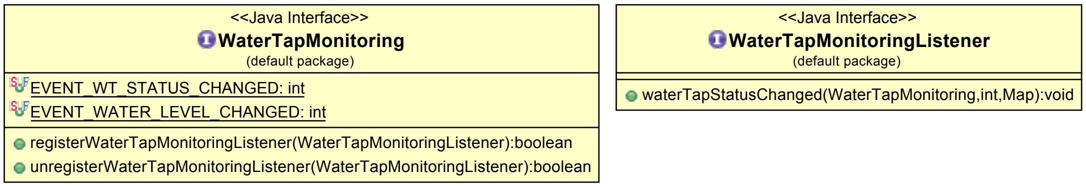


Figure 6.7: Water tap monitoring service: Class diagram.

of running water. Both features are covered by the proposed interface providing the following service events:

- `EVENT_WT_STATUS_CHANGED:int`
- `EVENT_WATER_LEVEL_CHANGED:int`

Registered listeners are informed about events raised by using `waterTapStatusChanged`. Each event provides a map containing optional information. Hence additional information such as the amount of water used since the last water level modification or the current time can be provided. Table 6.2 shows pre-defined map key-value pairs.

Table 6.2: Water Tap Monitoring: Pre-defined key-value pairs.

Key	Information	Format
time	event timestamp	String (yy-mm-dd_hh:mm:ss)
waterTap_id	water tap identification	String
waterTap_loc	water tap location	String
waterTap_status	water tap status	int
cons_water	consumed water (liter) since last event	double

In Chapter 3.4 a sound-based water measurement system was already introduced. However, it was also mentioned, that there are several water measurement systems already available based on different technologies. Consequently, service providers intending to develop functional services based on water tap monitoring services do not have to worry about the underlying technology (e.g. sound-based solutions or installed water meters between pipes). As long as the `WaterTapMonitoring` interface is realized, they are completely independent from the technology used. Thus, even end-users can use functional services without worrying about specific installations.

6.7.2.3 Ambient Monitoring – Electronic Device Safeguard

The third service is also related to ambient monitoring and provides an assistive service which is able to recognize electronic devices that have been left on. Such an application is for example of great interest for patients with dementia as many household appliances such as left on irons can cause dangerous situations. In addition to this fact, left on electronic devices waste a lot of energy. The electronic device safeguard service is based on power sensors as introduced in Chapter 3.5. The technological service considered has to register to *iSensor* devices and based on operating mode information and pre-defined maximum operating time thresholds the service is realized. Figure 6.8 gives a common interface description for an electronic device safeguard service.

Besides methods to register and unregister service listeners, an electronic device safeguard service provides the functionality to define maximum operating times (in seconds) for single *iSensor* devices. This is done using `setMaxOpTime4Device(iSensor,int):void`. If an electronic device was in use for a longer time period than the pre-defined threshold, the electronic

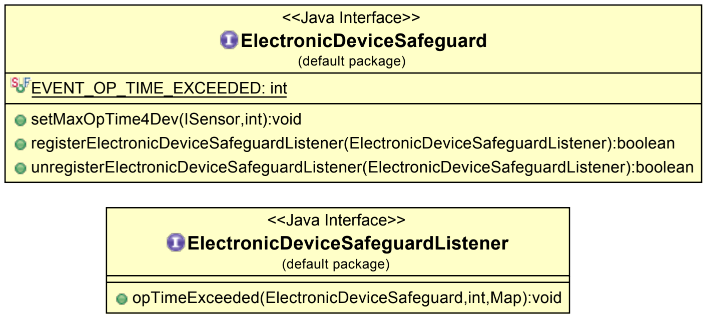


Figure 6.8: Electronic device safeguard service: Class diagram.

device safeguard service throws the event `EVENT_OP_TIME_EXCEEDED`. Registered listeners are informed using the method `opTimeExceeded`. Besides information about the service that throws the event, the event id and a map containing more detailed information is provided. So far features shown in Table 6.3 have been defined.

Table 6.3: Electronic Device Safeguard: Pre-defined key-value pairs.

Key	Information	Format
time	event timestamp	String (yy-mm-dd_hh:mm:ss)
iSensor_id	power sensor identification	String
iSensor_loc	power sensor location	String
iSensor_devDesc	description of connected device	String
op_mode	current operating mode	String
max_op_time	allowed operating time (sec)	int

Based on electronic device safeguard services, functional services can be realized that are able to automatically turn off left on devices or to inform the user. Consequently, by following the predefined interfaces high-level smart services can be realized without having detailed information about the used underlying technology or plugged hardware. So it is not important for software engineers if smart devices are used to submit their operating mode or if mainstream devices are turned into smart devices using plugged sensors. The only condition is, that devices and technical services have to follow standardized interfaces.

6.7.2.4 Communication – Handle SMS

The last example considers a service which is related to the communication group. In the case of detected emergencies such as fire, leaking gas pipes or user falls, smart services must be able to call for help. One possibility is to send an SMS to emergency departments or relatives. Figure 6.9 shows defined interfaces for a common service providing functionalities to handle SMS. Besides methods which allow other services to register and unregister to events, `sendSMS(String,String):int` must be realized. `String` variables are used to specify the target phone number as well as the message content. The SMS service informs registered listeners about the current state of a sent message. This means, that listeners are informed whether a SMS was delivered or sent. In the case of an error, a specific error description can be provided as `String`. If the HandleSMS service provides the functionality to receive SMS as well, incoming messages are forwarded to registered listeners using the `EVENT_SMS_RECEIVED:int` event. All events are provided to registered listeners using the `smsEVENT(HandleSMS,int,String):void` method. There, the calling `HandleSMS` service must be specified. Besides that, each event contains the corresponding event id as well as an optional event message (e.g. in the case of `EVENT_SMS_ERROR` a detailed error message).

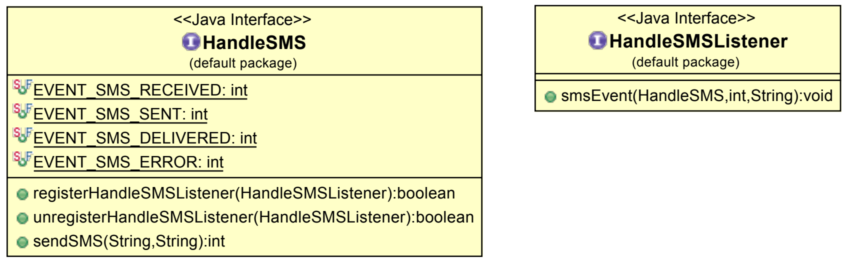


Figure 6.9: SMS service: Class diagram.

6.8 Development of AAL Services using OSGi4AMI

After a standardization approach and service interface examples were introduced, this section shows examples of functional services that are based on devices as well as technological services. In the following the two services $AppSURE$ and $ZoneSURE$, that were realized and installed in real end-user homes during the EU project MonAMI (www.monami.info) and are related to technology introduced in Chapter 2 and Chapter 3.5 are discussed. Consequently several project partners have been working on the realization of different smart home components. Because of this, already existing services developed by other partners (in detail: motion sensor device and ZigBee driver) could be easily re-used as they followed the standardization approach proposed.

6.8.1 AppSURE

AppSURE is a logical extension of the already introduced electronic device safeguard service. Figure 6.10 shows how AppSURE was designed. The service registers to available *iSensor* devices (handling the communication between real devices and the framework) as well as to the technological services $SMSHandler$ and $ElectronicDeviceSafeguard$. The last service will inform $AppSURE$ in the case of exceeded operation times of electronic devices. If such an event occurs, AppSURE is able to turn off the corresponding device by using the power cut functionality of an *iSensor* device. After that, a SMS is sent to the home owner in order to inform him about the deactivation event. The following rule describes the realization of $AppSURE$.

$$\textbf{IF} \quad EVENT\_OP\_TIME\_EXCEEDED \quad \textbf{THEN} \quad \begin{array}{r} iSensor.cutPower() \text{ AND} \\ HandleSMS.sendSMS(pn,msg) \end{array}$$

I want to highlight again, that the big advantage of using standardized interfaces is the fact, that software engineers do not have to consider the underlying technology. In this case for example, it does not matter what type of sensor is used to realize the communication with real electronic devices or to turn them into smart devices. This fact makes the development of functional services much more flexible and convenient.

6.8.2 ZoneSURE

$ZoneSURE$ is a smart service that was designed to monitor areas within homes. The motivation was that dementia patients are often not aware of the current time and hence they confuse night and day times. To prevent that people are active during night times, $ZoneSURE$ informs the nursing service by sending an SMS if persons are entering pre-defined areas within forbidden

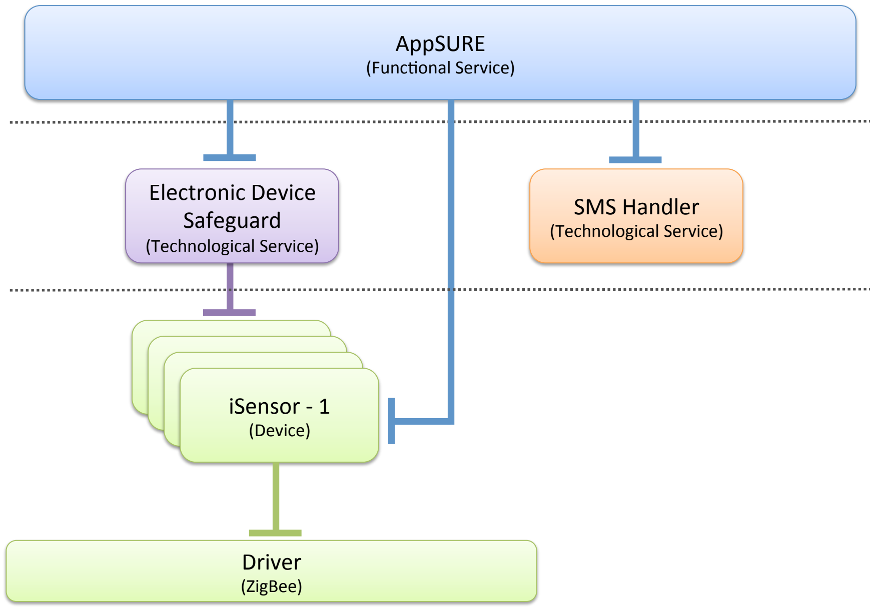


Figure 6.10: Realizing the functional service *AppSURE*. The service detects left on electronic devices, turns them off and informs the user about the deactivation by sending an SMS

time ranges. For example, persons are allowed to enter the toilet during night times, but they should not enter the living room or leave the house. Besides informing the nursing service, *ZoneSURE* can also be used to log user activities, to recognize behavior patterns and to track the course of diseases. Figure 6.11 shows the architecture of *ZoneSURE*. Based on technological services $UserLocationMonitoring$ and $SMSHandler$, *ZoneSURE* can easily be realized. This example shows again, that the development of *ZoneSURE* is completely independent from the underlying technology. Motion sensors as well as ceiling cameras are used to detect if a person is within a pre-defined area or not. As details about the implementation and the environmental setup are completely hidden, the realization of *ZoneSURE* can simply be described by:

IF (($EVENT\_AREA\_ENTERED$ OR $EVENT\_AREA\_LEFT$) AND $isForbiddenTime(AreaID)$) THEN $HandleSMS.sendSMS(pn,msg)$

6.8.3 AppSURE and ZoneSURE in Real-Life Large-Scale Scenarios

The MonAMI project realized all in all 32 assistive services based on OSGi4AMI. All services were installed in real-homes of end-users. There, only people over the age of 65 with at least one impairment in terms of vision, hearing, dexterity, mobility or cognition were considered. In summary 87 users from Spain, Slovakia and Sweden were chosen to evaluate functional services in their own homes for three months. In the following, evaluation results related to user experiences for *AppSURE* and *ZoneSURE* are shown. The evaluation was done by the London School of Economics (LSE)[75] and should illustrate the acceptance and the possibility of a real-world deployment of services based on OSGi4AMI.

AppSURE was installed in 16 houses located in Spain and Sweden. The end-user reports concerning *AppSURE* showed, that the only negative aspect was the size and the design of the sensor used. People mentioned that the sensor case is too large and bulky for real-life applications. Nevertheless users reported that they felt much safer since they used *AppSURE*.

[75] An overall evaluation description can be found at: http://www.hi.se/Sidanskatalog/14016/MonAMI%202011-09-26%20D34.3%20v2%20post%20review.pdf (last accessed on 2013/05/14).

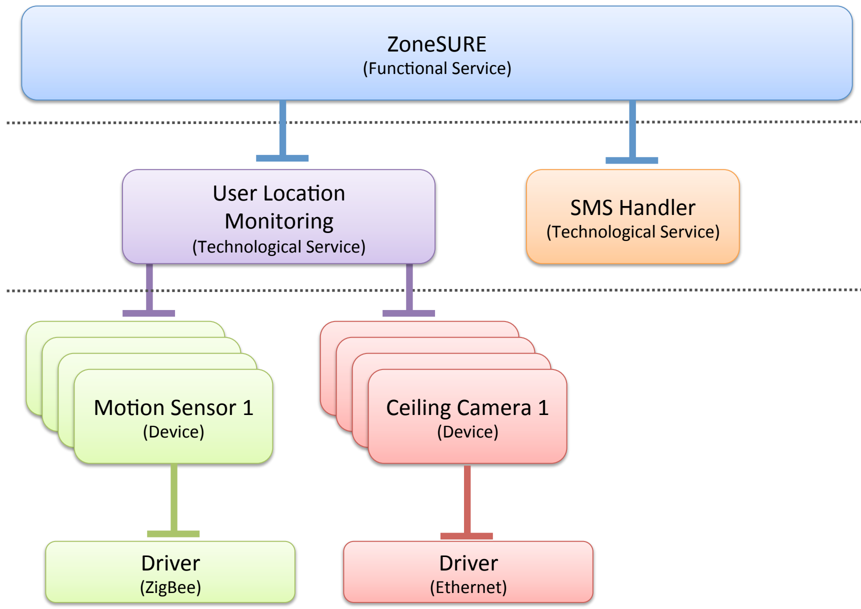


Figure 6.11: Realizing the functional service *ZoneSURE*. The service detects if people are moving within pre-defined regions during a specific time period. This information is used to inform nursing or security services by sending an SMS. In the case of dangerous situations, the police can be informed too. Besides that, *ZoneSURE* can even be used to recognize user behavior patterns.

The assurance that all electronic devices are turned off when leaving the home was a very important aspect for many users. Apart from that, the advantage of energy saving was also mentioned by many users.

ZoneSURE was installed in 38 houses located in Slovakia, Spain and Sweden. Negative reports regarded only the lack of more complex configuration possibilities. However, users in Slovakia rated *ZoneSURE* as the most helpful service in providing an autonomous life. Users mentioned many times, that they felt much safer when using *ZoneSURE* because of a significantly shorter response time to get help in the case of an emergency.

All in all the system was very well accepted. In the case of Swedish users, 71% reported that they were not impaired in daily life activities by sensors and services. This fact shows, that services introduced were unobtrusively integrated into existing environments. 32% even mentioned that services had influenced their ability to view the future optimistically as they know that intelligent services are there to prevent dangerous situations or to call for help immediately if needed.

This confirms again, that users are already open for smart services. The case study and MonAMI in general showed, that complex monitoring services can be designed distributed and based on exiting services when following standardized interfaces as was proposed in this work. Besides being able to save costs for elderly care, the fact that such systems are able to provide a more secure and independent life at home, shows the importance of making smart homes real.

6.9 Conclusion

This chapter introduced a standardization approach for services in smart home environments. The approach proposed was realized as an extension of OSGi. Service interfaces (OSGi4AMI) were introduced related to AAL applications. Besides a common introduction of smart homes and their components, OSGi was chosen as a suitable platform for residential gateways. Based on OSGi, components of AAL applications were defined grouping OSGi bundles in drivers, device/sensor representatives as well as technological and functional services. The definition of standardized interfaces for common components allows an easy and distributed development of new services as well as the re-usability of existing ones. In this work a standardization approach for technological services were considered. All in all more than 30 service types were addressed by OSGi4AMI.

As this thesis focuses on systems deployed in real world and large-scale applications, services were installed and used for several month in real homes of end-users. This fact shows, that the aim of distributed service development, the aspect of service re-usability and easy large-scale deployments in real-life scenarios can definitely be reached through the standardization approach proposed. If all service engineers would agree to follow a standardization approach as the one described in this work, the development of high-level functional services would no longer be restricted to software specialists. On the contrary, people without programming skills would be able to define their own high-level services by defining simple rules that combine already existing service modules. Consequently, the creation of services could even be done by end-users or people that are close to end-users such as nurses, caretakers or security companies. As such people are very close to real application fields, very specific and adapted services can be realized. This fact would result in a significant improvement in terms of service quality, service usage and service development. Besides that, it could open even a new AAL App market similar to the Apple AppStore or the Android Marketplace.

7 Thesis Conclusion

The recognition of human activities is one of the most highly addressed topics in pervasive computing. It has already been investigated, that the location of people indoors as well as information about device usage can contribute to a better performance of activity recognition systems. However, as was already stated at the beginning of this thesis, many systems are based on complex models, expensive sensors, highly instrumented environments or large amounts of training data. Consequently, real-world and large scale deployments are complex, time consuming and cannot be done by people with non technical background in many cases.

The main part of the thesis works towards the solution of this problem. It was shown, that issues such as indoor positioning or the recognition of specific device operating modes can even be reliably solved when using approaches based on minimal training data sets, easy configurable models, minimal invasive instrumentations and low-cost sensors. The main focus was on a difficult activity spotting problem including subtle, barely distinguishable hand activities. Before this problem was addressed, solutions were presented which can significantly contribute to a better recognition performance of the spotting problem considered.

The thesis starts by introducing a sub-room level indoor location system (see Chapter 2). Methods were presented, that were able to solve the issue of user identification in simple computer vision tracking systems through sensor fusion. In Chapter 3, the recognition of daily-life activities were addressed. The thesis showed, that besides the fact, that common household devices were used, approaches proposed are even able to detect *how* and *what* appliances have been used *for*.

As a next step, it was shown how concepts and systems introduced can be used to improve the performance of a wearable system in a difficult activity spotting scenario. Apart from that, the thesis presented how and to which degree various systems and their fusions, targeting search space restrictions, could contribute to a improvement of performance. Problems, which occurred due to the fact, that minimal training data and easily configurable models were used instead of complex systems relying on large amounts of training data, were compensated by using sensor fusion approaches (see Chapter 4).

Furthermore, the thesis presented an unobtrusive solution for the problem of detecting stress situations of individuals. While state-of-the-art approaches use physiological parameters and obtrusive skin-attached sensors, the thesis focused on the detection of significant user behavior changes due to mid-term stress situations. In general, behavior monitoring systems are based on the surveillance of basic user activities as were introduced in Chapter 2 – Chapter 4. However, the monitoring of stress situations cannot be reliably solved on the basis of such indoor activities alone. A concept was introduced, that was able to solve this problem based on smartphone integrated sensors. Consequently, the system is completely unobtrusive and environment independent (see Chapter 5).

Finally, activity recognition and behavior analysis systems must be integrated into smart home gateways if they should be deployed in real environments on a large scale. Hence, the closing part of the thesis presented a standardization approach for the realization and integration of pervasive computing services (as were shown in this work) in a mainstream residential gateway framework (see Chapter 6). Medium-scale, real-world deployments and evaluations of ambient intelligence services based on the standardization approach introduced have illustrated its usability and user acceptance in such scenarios. Therefore, realized services were developed based on concepts introduced in Chapter 2 and Chapter 3.

In the following, key contributions achieved and the findings of the thesis are summarized. Furthermore an outlook is given for further research questions, that have emerged during the investigations performed. Finally, practical implications and questions are discussed.

7.1 Summary of Contributions

The overall challenge of this work was to show that models based on physical constraints and simple one-time measurements can be used to solve the problems considered instead of using complex configurations, difficult deployments or large amounts of training data. The key contributions of each considered topic achieved are shown in the following.

Indoor Location: While most indoor location systems are based on expensive sensors and highly instrumented environments, Chapter 2 presented a low-cost and easy-to-deploy indoor location system providing sub-room level accuracy. The system is based on a simple and standard computer vision foreground-background detection and motion tracking approach. Although such systems are unobtrusive and easy to set up, they cause the following problems:

(a) They are unable to identify tracked objects. Only unidentified moving objects are detected and tracked.

(b) Due to the fact that they were designed to detect movement in general, issues related to small movements (e.g. hand movements), lost objects as well as wrong object assignments in indoor, human tracking scenarios occurred.

(c) They only provide pixel coordinates for detected unidentified objects, which have to be mapped to real-world locations.

The thesis has overcome these problems. First, easy to parametrize methods and models based on physical conditions were introduced in order to solve the problems described in (b) and to adapt standard computer vision motion tracking system to the needs of an indoor, human tracking scenario. Secondly, the thesis presented a correlation approach, that combines simple motion patterns derived from low-cost on-body motion sensors and unidentified, tracked objects to solve the identification issue described in (a). Finally, pixel-based object locations were mapped to pre-defined polygonal lines (regions of interest) in order to map pixel coordinates to real-world, sub-room locations (see (c)). Additionally, the issue of transmitting personal data, which raises privacy concerns, was addressed and weakened. Although the scale of the experiments is small, they demonstrated, that fusing simple, easy-to-deploy systems can solve a challenging identification problem instead of using complex face recognition systems or expensive, difficult to deploy high precision trackers.

Use-Mode Recognition of Common Household Appliances: Detailed information about specific use-modes can contribute much more to systems targeting at complex activity recognition problems or behavior analysis than simple "device in usage" information. While a large amount of state-of-the-art approaches addressed the problem of recognizing *which* household device was used, the thesis showed solutions for the so far unconsidered problem of identifying *how* and *what* they were used *for*. The thesis introduced easy to deploy and easy to maintain solutions for water taps and electronic household appliances. Up to now, these devices have been frequently addressed by state-of-the-art approaches.

First, the detection of use-modes of common water taps was investigated. The following problems occurred due to a real-world, large-scale application scenario:

- Operating modes of common water taps mainly include the following aspects: the detection of flowing water, the recognition of a specific water flow level and the approximation of the amount of water consumed. State-of-the-art approaches have mainly used audio analysis to detect water tap usage events in general. Based on this fact, the thesis investigated how well different water flow levels can be detected and how well the amount of water consumed can be approximated by using low-cost and easy to deploy microphones.

- In order to enable large-scale deployments, the focus should be on a minimal data set recorded by simple, one-time measurements.

- Surrounding noise is present in almost all real-world environments. As the recording of all environmental sounds is difficult as well as time-consuming, a solution must be found to filter out such noises and/or to reduce the impact of false classifications without reference data.

A minimal data set was recorded for several discrete water flow levels. Based on that, audio processing features were defined and combined with several classifier paradigms. It turned out, that a kNN classifier is able to distinguish between several water flow levels very well and furthermore is able to outperform a decision tree classifier as well as Support Vector Machines. Clearly, the calculation of the amount of water consumed provides much more valuable information for activity recognition applications than discrete water flow levels. The thesis solved this problem using calculations based on physical concepts and information about the current water flow. Finally, it was shown that rules related to logical boundary conditions are able to solve the problem of detecting surrounding noise and to reduce the impact of false classifications as well as outliers. Several real-life evaluations have proven that even when using minimal reference data, water flows can be reliably distinguished by the systems proposed.

Secondly, electronic devices were considered. While state-of-the-art approaches analyze transient noise to identify devices being used, the thesis introduces methods to recognize specific use-modes on top of identifying the devices. The following problems occurred due to a real-world, large-scale application scenario:

- In order to guarantee an out-of-the-box solution, data processing and classifications must be performed on the sensor device itself. Consequently, related features and classifier paradigms must be realizable on low-cost, low-power CPUs such as those used in embedded systems.

- In order to reduce the amount of sensors necessary, methods must be defined to handle several simultaneously operated devices at the same time.

Features, that are easy to calculate and methods based on electric current values were introduced. An in-depth signal analysis showed, that the defined features and rule classifiers are able to recognize device use-modes when devices are being operated independently. Furthermore, the analysis of signals generated from simultaneously operated devices were investigated. It turns out, the defined methods and features are powerful enough to solve the problem of detecting operating modes of multiple devices with one single sensor. However, it was shown that due to the sensor design, the simultaneous usage of high power consuming devices is quite restricted. Furthermore, it was investigated if detected use-modes can be described in even greater detail. This idea was based on the consideration that the power consumption for the same operating mode depends on what the device was actually been used for. It is shown, that the proposed methods and features are also able to solve this problem. For example, it was shown that mixing something liquid is less power consuming than mixing something creamy and hence it was possible to recognize the rigidity of a mixed liquid as well as the process of the rigidity change.

Spotting of subtle, barely distinguishable hand activities: The spotting of user activities and object interactions embedded in a large amount of background data is one of the most challenging problems in activity recognition. State-of-the-art approaches, that were successfully applied to similar problems, are widely based on very large data sets representing the targeted problem and consist of several repetitions of each activity. In contrast, this thesis solved the problem by considering minimal data sets and simple to perform one-time measurements and configurations. To that end, models describing commonly valid activity pre-conditions and unobtrusive, easy to deploy approaches were introduced.

As a first step working towards the recognition of subtle hand motions, the spotting of related object interactions was focused on. A multi-modal core system (wrist-mounted camera, orientation sensors and IR-proximity sensor) was introduced which was able to significantly outperform a state-of-the-art inertial sensor approach. However, the achieved results were not sufficient for many real-life applications.

Consequently, in a second step, it was investigated how and to what degree the performance of the core system can be improved by further sensor fusion. The key idea was, that the search space of the core system can be restricted further which leads to a significantly increased system precision. State-of-the-art approaches have already proven that hand movement features can contribute to the solution of similar problems. Hence, easy configurable models related to hand movements as well as time features were introduced and investigated. Besides that, location and operating mode information can clearly contribute to the solution of such problems as they are able to provide meaningful information about the current situation of the user and of a device. The thesis has already introduced related concepts in terms of minimal deployment effort in Chapter 2 and Chapter 3. It was shown, that fusion approaches based on these systems have achieved the most significant improvements. Finally, the impact of fusing multiple sensors was investigated. Again, it turned out, that a combination of location and operating mode information achieves the best results.

However, the spotting of object interactions does not provide information about concrete user hand activities, which are the focus of this work. Consequently, in a next step, it was investigated how existing modalities can be extended to recognize even subtle hand activities. The thesis analyzed if inertial sensors and device operating mode information can be used on top of previously spotted object interactions in order to recognize underlying subtle arm activities. Although it was already shown, that inertial sensors and representative data sets are not able to provide sufficient results to spot subtle activities, they were used to distinguish between a finite set of object related subtle hand motions. It turned out, that inertial systems were not able to provide sufficient results for this problem again. Furthermore, it was shown, that mapping operating mode information to arm activities can significantly outperform inertial systems. The thesis also investigated the combination of both systems in order to compensate for their respective disadvantages (e.g. whereas it is almost impossible to use only motion sensors to differentiate between a button being pressed to turn a device on or off, this issue can be easily solved by monitoring the use-mode of the device). Unfortunately, the results were worse than those resulting from fusing only information about device operating modes. However, the number of recognizable subtle hand activities was maximized.

Finally, due to the fact, that the main component of the core system is based on a computer vision object recognition approach, which is known to be computationally quite expensive, the thesis investigated, if the camera component of the core system can be replaced by an inertial system. Furthermore, it was analyzed if the said system could contribute to an improved performance of the core system. It was shown, that inertial sensors are not able to contribute to a performance improvement of the core system at all. Apart from that, the replacement of the camera sensors results in a significantly worse performance and therefore makes the system unusable for many application scenarios.

Detection of User Behavior Changes due to Medium-Term Stress Periods: Due to the fact, that nowadays the burnout syndrome has become a widespread disease, many research efforts have been dealing with the detection of stress situations. State-of-the-art approaches are based on the monitoring of physiological parameters and use obtrusive skin-attached sensors or voice analysis. In contrast to that, this thesis presents a completely unobtrusive way to detect significant behavior changes due to medium-term stress situations.

The concepts and systems introduced in Chapter 2 – Chapter 4 provide concrete information about user activities, their location as well as how devices were used. They can therefore be used as basic components for systems learning behavioral patterns and detecting significant behavioral changes. However, to reliably detect stress situations, a 24 hour monitoring must be guaranteed. Clearly, the possibilities of the systems shown are quite restricted in such scenarios as they rely on environmental instrumentation - even if it is minimal. The thesis presented a concept able to derive behavioral parameters from individuals based on smartphone integrated sensors only. Consequently, it introduced an unobtrusive solution for long-term monitoring of stress levels. The following problems were addressed:

- Which behavioral parameters, that are affected by medium-term stress situations and are based on smartphone sensors can be defined?
- How can a real-life data set be recorded that consists of both a medium-term stress-free and a medium-term stressful period?
- If smartphones are used to monitor behavioral parameters, it has to be guaranteed, that the devices can still operate as normal smartphones in terms of functionality and battery life.

First, the thesis defined behavioral parameters related to the user's location, social interaction behavior and his/her phone usage behavior. Location features were defined as it is obvious that people might visit different locations depending on the stress level (e.g. spending more time at work, having less time to spend in a the gym, etc.). Furthermore, several approaches have already proven, that crowd estimates and physical meetings with people can be derived by smartphones. Stress can also have an impact on the kind of social behavior (e.g. the person has less time to spend in crowded environments such as shopping malls or cinemas). Consequently, those features have been considered as well. Finally, the analysis of phone calls and sent/received text messages has already been used in many state-of-the-art approaches to determine friendship between people. The idea is, that such features might also be useful to detect stress situations as people may try to compensate for a decline in physical meetings due to stress (e.g. with friends).

Secondly, a real-life data set was recorded. In order to get a realistic scenario with continuous stress situations, students were recorded during and after their exam period. In order to guarantee a certain stress level, only students were chosen who ranked exams as important but quite unrealistic to pass. The result of this recording was a dataset containing more than 40 GB of data.

Third, the thesis showed that the defined parameters were able to reflect medium-term stress situations of almost all participants. However, it turns out that the impact of those features is quite user-dependent and therefore commonly valid features can hardly be found. Besides that, the following problems related to the calculation of behavioral features were considered:

- Battery life: In order to guarantee a 24/7 monitoring, the sampling rates of power-consuming sensors such as GPS were reduced. It turns out, that even low sampling rates (a GPS fix was recorded every 10 minutes) provides enough information to determine the approximate retention time of a person at specific locations.
- Smartphone usability: Smartphones still need to be usable while behavioral features are calculated in the background. This fact is quite important, as otherwise people would not carry their phones with them all day and consequently meaningful behavioral features could not be derived from smartphone sensors.

OSGi4AMI - A Standardization Approach for Services in AAL Scenarios: The closing part of the thesis proposed a solution for the problem of missing standardization approaches for pervasive computing systems. Systems, as were introduced in this thesis, must be integrated into existing smart home gateways in a standardized way if large-scale scenarios are to be realized. Due to missing standardizations, service providers are still using proprietary protocols and provide closed systems. Consequently, the re-usability of services, distributed service development and large-scale deployments are suffering from this fact.

The following problems were addressed:

- To what degree can standard interfaces be defined without restricting the creativity of service developers?
- Which common service types can be defined in ambient assisted living applications?
- How can standardized interfaces be easily integrated into common gateway frameworks?

First, common service components were defined. While state-of-the-art approaches have largely proposed a strict distinction between devices and applications, this thesis introduced a solution considering the architecture of common services and defined three main service components: devices, technological services and end-user functional services. Based on that, interfaces for technological services (providing context and activity recognition information) were introduced by the thesis. More than 30 interfaces for common technological services related to ambient monitoring, personal support or communication were defined. Amongst others, the concepts already shown in Chapter 2 and Chapter 3 were considered. In order to allow service providers enough creativity, technological services provide information through containers and pre-defined key-value pairs. Hence, if services provide exotic or unusual information, developers can extend existing containers by new key-value pairs. Device interfaces are not addressed in this work as they were realized by Alvaro Marco and Roberto Casas (University of Zaragoza, Spain). End-user services have not been standardized due to the fact that they are wide-ranging and therefore the creativity of service developers would suffer too much by a standardization approach.

Secondly, the thesis has addressed the topic of choosing the optimal residential gateway platform, used to integrate the interfaces presented. The discussion shown confirmed the widely accepted usage of OSGi due to advantages related to its modular structure, available methods to manage services and the large amount of existing extensions related to various application scenarios. Because of these reasons, the interfaces presented were provided as OSGi extensions. Consequently, already existing methods to update and manage services remotely could be adopted.

Thirdly, the thesis showed, that two services related to ambient assistive living applications were realized based on the interfaces presented and by re-using existing components. Services (related to sub-room level monitoring and device operating mode surveillance) were deployed for four months in a total of 54 homes inhabited by elderly and disabled people. Besides that, it is worth noting, that the installation procedure was carried out by people with non technical background. Evaluations, performed by the London School of Economics, showed that the objectives of the approach proposed were fulfilled. Services were integrated unobtrusively into existing homes and end-users stated, that the systems had a positive influence on their daily lives.

7.2 Outlook and Open Questions

During the investigations made, several aspects and open questions for further analysis occurred. These aspects are summarized in the following.

Indoor Location: It was shown, that simple computer vision-based foreground-background detection systems do not perfectly fulfill the requirements of tracking humans indoors based on ceiling cameras. Although some issues related to this topic could be solved, a reliable tracking of people walking close to each other or in crowded environments in general are still open problems. In order to reduce the impact of resulting false classifications, wrong assignments must be detected as fast as possible. A feasible solution would be to focus on more detailed movement information. Besides basic modes of locomotion, between different discrete movement speeds like running or walking slowly could be distinguished. Due to the fact, that today's smartphones even include gyroscopes and magnetometers, even features like turning left/turning right or the walking direction can be taken into account. Obviously, this would result in a more powerful correlation process. The idea of considering the walking direction to assign unidentified tracked persons to on-body motion sensors was presented by [TJS10] two years after the approach shown in this thesis.

Use-Mode Recognition of Common Household Appliances: With respect to ***water taps***, it was shown, that an audio based system is able to approximate the amount of water used quite well (10% error) by taking into account only a restricted amount of discrete water flows (six levels of flowing water plus silence). However, the following unsolved problems were investigated.

- First, during real-life evaluations it turns out, that short water tap events (e.g. sketchy cleaning of silverware or washing hands) are not recognized by the system. Although the amount of water used is low, it could result in a considerable error during common cooking events. To overcome this problem, the rules introduced can be adjusted. However, the amount of false classifications will benefit from this modification and hence a trade-off must be found for each specific application scenario.
- Secondly, the thesis showed, that core system parameters are environment (water tap) independent. This fact leads to a facilitated deployment process. However, reference data for several water flow sounds and silence still has to be recorded in order to train the system. Although the training procedure is simple and easy to perform, the system is not completely environment independent.

The idea to monitor water taps and to recognize activities based on water tap usage events by sound analysis were investigated by many authors during the last few years. However, the work described was among the first publications in this field. Besides that, it has received a "Best Commercial Potential" award, which confirms, that the requirements related to easy deployment and unobtrusive instrumentation were completely fulfilled. The importance of monitoring water taps and water usage is emphasized by on-going research work. For example, [TSGM10] combined computer vision and audio features to detect water flows in hand-washing tasks. In contrast, water waste was detected by [VSN+11] using audio features.

With respect to ***electronic devices***, this thesis presented a system that is easy to extend and deploy and which has been able to recognize device use-modes. Besides that, even the fact *what* devices have been used *for* could be recognized in some cases. However, the following problems were investigated:

- First, the presented sensor is not able to handle high voltage current. Consequently, it could not be evaluated if the defined features and rules are able to recognize use-modes in such scenarios. Due to the fact, that common ovens require high voltage current and that they provide meaningful information for daily life user activities, a solution must be found to integrate even such devices.
- Secondly, the system will fail if different devices showing almost the same power profile are operated using the same sensor. Consequently, an additional way to identify the device must be found. Amongst other possibilities, RFID based solutions or a system as described in Chapter 4 can be used to solve this issue.
- Thirdly, the thesis showed, that simultaneously operated high power consuming devices can only be distinguished to a certain degree due to the design of the sensor used. Especially in kitchen scenarios such situations may appear quite often as related devices like water boiler or mixer are very power consuming. Therefore, the sensor has to be revised in order to handle this problem.

- Fourthly, the system works only for wired devices so far. However, many common electronic devices are wireless (e.g. electric tooth brush or shaver). Hence, operating modes of such devices can only be recognized by monitoring corresponding charger devices.

The importance of electric current sensing is emphasized by on-going research work and the large amount of approaches presented aiming at the recognition of electronic devices operated as well as high-level context information. I would like to note, that the work presented was among the first publications in this field. It received a "Best Paper Award" and was cited by 22 research papers[76] during the last few years.

Spotting of Subtle, Barely Distinguishable Hand Activities: The thesis has shown, that even easy configurable models and systems based on minimal training data sets are able to solve difficult activity spotting problems. However, the following problems were investigated:

- First, the height of objects is a key information for the presented core system. It is compared with the user's vertical hand height in order to select a set of feasible object candidates. Consequently, the system will have problems in scenarios in which objects are frequently re-placed (e.g. forks or tables). In such scenarios, an unacceptable reconfiguration process has to be performed to update the vertical position of objects.

- Secondly, one of the core components of the proposed system is a wearable camera. Consequently, the system has to face well-known problems related to computer vision applications such as low light or object occlusions. Nevertheless, the thesis showed that the camera is a key component of the proposed system. We have seen, that the core system is not able to achieve similar results when replacing the camera with an inertial approach. In contrast, significantly worse results were reached which make the system unusable for all application scenarios.

- Thirdly, the thesis has shown, that sufficient results can only be reached if the device operating modes are taken into account. However, many devices cannot be turned into smart devices by the systems presented in Chapter 2 and Chapter 3. Examples are forks, plates or cupboards. To overcome this problem, other solutions must be found. Due to the fact, that many "operating modes" of objects (besides electronic devices) are related to motion (e.g. door/cupboard opened/closed, fork used or cup filled with water), acceleration sensors can be attached to such devices in order to derive basic, unidentified use-modes ("object has been used somehow") for example.

- Fourthly, the thesis has shown, that the proposed system is able to reach sufficient results but is dependent on environmental instrumentations - even if they are minimal and unobtrusive. An obvious improvement would be to replace environment dependent approaches by on-body systems. We have already seen, that common inertial sensor approaches cannot contribute at all to a improvement of performance. The reason is, that only subtle and barely distinguishable activities were considered and hence characteristic motion features are missing. A promising idea is to analyze characteristic fragments of a subtle hand activity instead of considering the whole activity signal. This idea has already been mentioned in [KVLL13]. Another promising idea would be to integrate capacity sensors built in wristbands as shown in [CBL12]. In contrast to motion patterns, this system is able to recognize processes which occur under the skin. Such features may provide meaningful information, even for subtle and barely distinguishable hand activities.

- Fifthly, the computer vision based object recognition algorithm may be improved by using more powerful SVM kernels such as a RBF or a histogram intersection kernel (e.g. [BOV03]). Besides that, the SVM could be replaced with Joint Boosting methods which have been widely used in computer vision applications (e.g. [MF09]). Furthermore dependencies between activities and object interactions could be defined and integrated into

[76] arnetminer.org (last accessed on 2013/09/19).

the system based on context-free grammars (e.g. [RA06]). Finally, the inertial based approach and hence the system's ability to identify hand activities based on recognized object interactions may be improved by considering more complex features and methods such as the longest common subsequence (e.g. [NDRCT12]) or derivative dynamic time warping (e.g. [BD13]).

Detection of User Behavior Changes due to Medium-Term Stress Periods: The thesis has introduced a concept based on behavioral parameters which is able to detect medium-term stress situations in a completely unobtrusive way. However, during the investigations, the following problems were identified:

- Firstly, it is clear, that the amount of participants and the experiment's duration is not big enough to prove reliably achieved results. Consequently, an extended evaluation should be performed to confirm results reached so far.

- Secondly, the system presented behavioral parameters based on location, social interaction and phone usage. However, smartphones provide many more possibilities. Hence, behavior features related to activity patterns, phone application usage or sound analysis should also be integrated and investigated.

- Thirdly, so far only single behavioral parameters have been considered. Hence, future work should also investigate multiple parameter combinations and their correlations.

- Fourthly, although the fact, that students have to face significant stress situations during their exam period is obvious, a more reliable way to get ground truth information must be found. A possible solution would be to use self-report questionnaires as they have already been used in related approaches (e.g. [JPS+13]).

Finally, it should be investigated if the combination of behavioral and physiological features can improve the reliability of stress detection systems. A first approach is shown in [GPT+]. There, acceleration sensors were used to get information about current user activity. Based on this information, ECG signals can be evaluated more reliably as they are strongly influenced by physical activities.

OSGi4AMI - A Standardization Approach for Services in AAL Scenarios: The closing chapter of the thesis introduced a standardization approach for pervasive computing services in ambient assisted living scenarios. Although we have seen, that the interfaces proposed and their integration into OSGi are definitely a suitable solution for the problem considered, the following problems were investigated:

- First, containers with pre-defined key-value pairs were used to provide recognition results or high-level information. Hence, service providers are not too restricted in their creativity to design services as these definitions can be easily extended. However, the idea of standardization would be destroyed if service providers prefer to define their own access parameters instead of using pre-defined values.

- Secondly, so far the proposed standardization approach has been designed for and integrated into OSGi. Although OSGi is widely used as a residential gateway platform, it is not "the only" solution. Hence, a way must be found to provide the proposed standardization approach as a stand-alone solution that can easily be integrated into different gateway platforms.

So far, end-user functional services have not been standardized. The reason is, that these services are wide-spread and hence service providers would be too restricted in their creativity to design innovative services. Although end-user services are mainly based on linking technical service components, the realization of new functional services still requires in-depth programming knowledge. If a way can be found to create such services by simple "drag and drop" solutions, the development of functional services could even be done by the end-users themselves. Thus, a new market similar to the Android Marketplace or the Apple App Store could be born. Obviously, the required basics for such an approach are standardized service components such as were introduced in this thesis.

7.3 Practical Implications

As was already stated in the beginning, three of the most important requirements for real-life and large scale service deployments are that services offered are based on minimal training data, unobtrusive environmental instrumentations and low-cost sensors. The realization of activity and context recognition services keeping these conditions was the main focus of this thesis. But in addition to the implementation of such services and their standardized integration in smart homes, there are still several open issues that have to be solved in order to bring smart services onto the mass market. Examples include various topics such as privacy issues related to the processing of personal data, security problems or the design of user interfaces. Some of them have already been addressed in several publications or EU projects as MonAMI. In the following, three issues related to service management, Human-Machine interactions and the evaluation of service usage statistics, which were addressed by MonAMI, are discussed. In general the MonAMI system provides a feasible solution for smart home environments developed for large scale scenarios and adapted to the special needs of elderly and disabled people. The main idea is described in [FKW+10]. A standard touch screen PC was used as residential gateway (RG) running OSGi and the introduced extension OSGi4AMI. Sensors and actuators were integrated in the MonAMI system by using standardized interfaces defined in OSGi4AMI.

7.3.1 Service Installation, Configuration and Maintenance

As a first step, services must be made available to end-users. The ideal solution might be to provide a central AAL service App store (similar to the Apple or Google App Store) where assistive services can be purchased, downloaded and updated. MonAMI provided a solution which allows service providers (which were represented by the telco providers France Telecom and Telefonica ID) to upload services to end-user gateways and to adapt them to the needs of end-users. Therefore each service had to implement specific interfaces which were also included in OSGi4AMI. Consequently, telco providers were able to adapt installed services to the special needs of end-users or even service them. The configuration was done via web interfaces. Besides general system settings such as language or user information, each service is configurable. So the MonAMI system makes it possible to specify how and who should be contacted in the case of a service event. Three notification channels have been considered so far: email, SMS or phone call. Services can also be enabled and disabled. In this way, users or caregivers can decide how long and at what times services should be active. Hence, the system is more flexible and able to respect privacy issues. One example could be that people want to be monitored only if they are alone at home. In such scenarios husband or wife can handle daily life issues such as going to the hairdresser while their partner remains secure, alone at home.

7.3.2 Human – Machine Interaction

Due to the fact that systems must be able to handle different user disabilities and impairments, various input devices and user interfaces have to be considered by a smart home system. Hence a common way must be defined to integrate several input / output devices in the system. In MonAMI this issue is solved by splitting the Human-Machine interface (HMI) block from services running OSGi. So, the HMI runs on a Universal Control Hub (UCH), which is a HMI server implementing the ANSI standard Universal Remote Console (URC)[77]. In this way the development of HMI clients on different target devices such as tablet PCs, smartphones or

[77] http://myurc.org (last accessed on 2013/06/10).

exotic devices was realized which are adapted to the special needs of their users. Finally the user interface itself can be changed and adapted to end-user needs and roles. Caregivers for example can operate with user interfaces providing much more functions than the elderly who need quite clear and easily structured interfaces. MonAMI provided several standard UI layouts and the ability to integrate new ones with ease. Figure 7.1 shows a MonAMI user interface which has been adapted to touch panels.

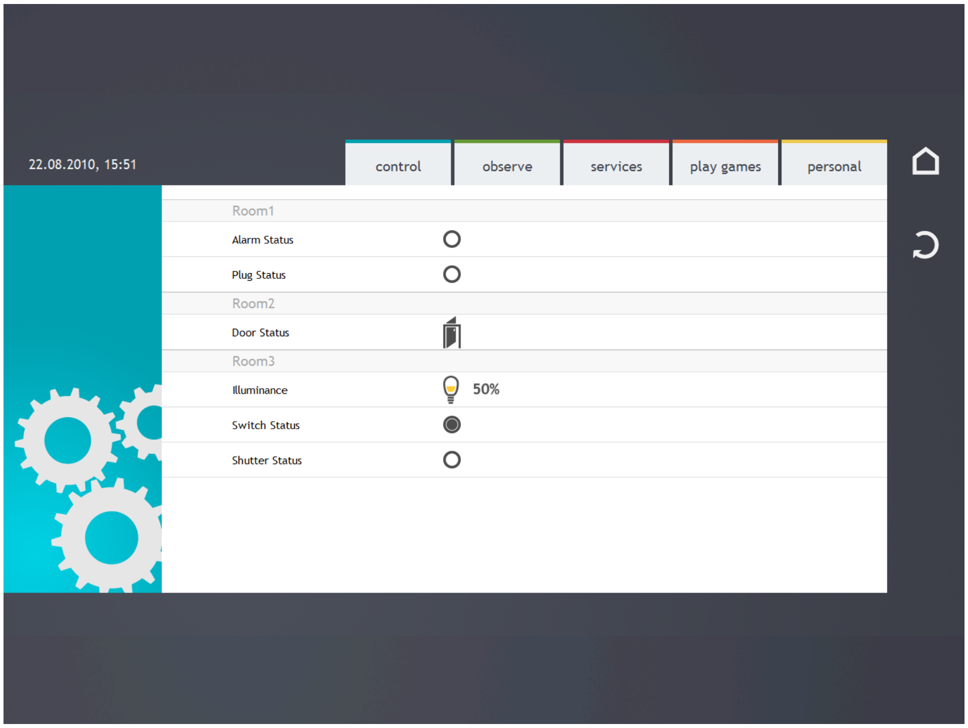


Figure 7.1: MonAMI User interface: Service Control

7.3.3 Service Usage Statistics

Finally, a quite important issue for service providers or caregivers is the analysis of service usage statistics. The amount of downloaded or installed services can easily be determined as services are handled by provider operated platforms. In contrast to that, logging how often a service was finally used by end-users or how often services have interfered in the daily life of end-users is much more difficult. This information could be very interesting in terms of creating service rankings or introducing new business models. The problem is that information about service usage must be transmitted from smart homes to telco platforms. MonAMI handled this problem by transmitting system log files. Each service was able to write log information by implementing a simple logger interface which was included in OSGi4AMI. A functional service realizing logger functionalities received usage information from all services, buffered and transmitted them to service providers or caregivers by email. As it is obvious, that evaluating human readable log files by hand is very time-consuming, I supervised the development of a log evaluation tool. The tool is able to read log files and to visualize service usage statistics for several days and for specific services. Figure 7.2 shows a screenshot of the system. Log files can be loaded and entries can be displayed and compared with each other on a three layer circle. Each circle shows entries for one day and arranged by hours. Bubbles represent the amount of service entries by their size. The tool also provides the possibility to zoom in and to show service usage information in more detail (e.g. on the basis of seconds). Once a single entry is selected, detailed information about service usage time, event location and event specific messages are displayed. Apart of

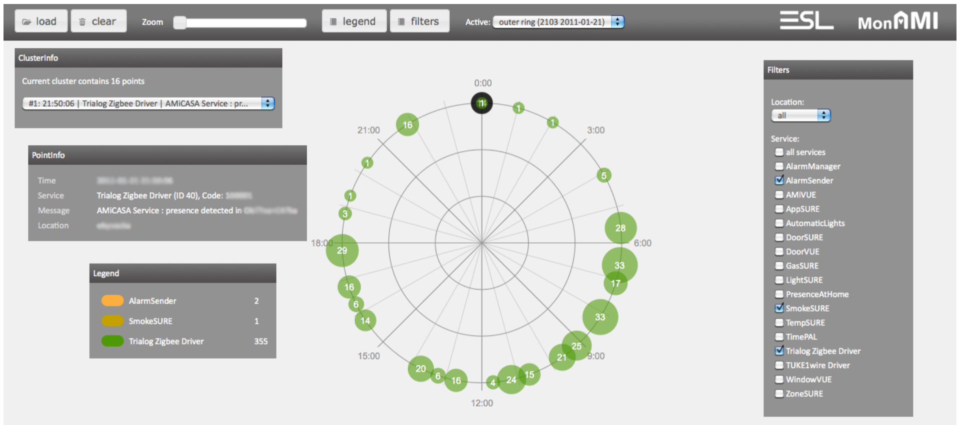


Figure 7.2: MonAMI Log Evaluation Tool.

that, displayed services can be filtered by location and service types. All in all the tool provided powerful functionalities to analyze and visualize service usage statistics. Of course the tool is not "the" solution to evaluate common log files. However, it was a suitable solution for MonAMI which was very well accepted by both service developers and end-user representatives.

At this point, the statement, that was made in the introduction, is being revisited. Are services as were described in this thesis and the systems which were realized by MonAMI really the solution of the stated problem? Will assistive services and smart homes conquer the mass market as soon as the aspect of *Applicability* has exceeded a certain level? Of course nobody can guarantee this fact. But even if they do not find their way onto the mass market, the increased usability of services will open the way for integrating smart systems and assistive services into more application scenarios such as public buildings or nursing homes. It is clear, that our society will have to care for more and more elderly and disabled people in the near future. Even today, nursing homes or hospitals are filled to overflowing and the nursing staff is unable to cope to a large part. Assistive services and smart environments as they were described in this thesis can definitely increase people's quality of life. Such systems make it possible for elderly and disabled people to handle their daily lives independently for a much longer time and consequently nursing homes and caregivers can be relieved. Of course many more aspects than covered by this thesis have to be considered in order to realize a smart home system which is ready for the market. But quite certainly, when it comes to service realization the most important aspects will be: easy system setup (minimal training effort), unobtrusive integration and low-cost sensors.

Abbreviations

This section gives an overview about important definitions, variables and abbreviations which were used in this thesis. All declarations are grouped by chapters and listed in an alphabetical order.

Design of a Sub-Room Level Indoor Location System

$accDev$	Motion sensor.
$accDev_{act_i}$	The i-th movement pattern of a motion sensor.
$accDev_{act}$	A motion sensors' mode of locomotion patterns.
act	Motion type (Standing / Walking).
$Blob$	Moving objects recognized by computer vision techniques.
$blob_{act_i}$	The i-th motion pattern of a blob.
$blob_{act}$	A blobs' mode of locomotion patterns.
$blob_{id}$	Unique blob identifier.
$blobHandler$	SM extension to overcome issues related to small movements, wrong blob allotment and lost blobs.
m	Distance threshold for merging new unassigned blobs in immediate vicinity of existing blobs (unit: pixel).
n	Threshold on blob height and width (unit: pixel).
ROI	Region of Interest.
s	Standing pattern.
SM	OpenCV surveillance module.
t	Timestamp.
thr_{time}	Maximum allowed time deviation between a motion sensors' and a blobs' movement change (unit: seconds).
w	Walking pattern.

Operating Mode Recognition of Mainstream Household Appliances

ADC	12-bit analog - digital converter.
ADC_{peak}	ADC peak calculation based on the last 600 raw ADC values.
$duration$	The time duration in which a device was operated in a specific use mode (unit: seconds).
$iSensor$	Sensor used to calculate the current consumption of connected electronic devices based on electromagnetic induction.
kNN_{dist}	Sum of distances between a measured feature vector and the k nearest feature vectors of the current winner class. This threshold is used to distinguish sound samples of flowing water from background noise.
kNN_{NR}	Water flow measurement system based on a kNN classifier and a noise reduction layer.
kNN_{NR+RS}	Water flow measurement system based on a kNN classifier, a noise reduction and an additional rule layer.
kNN_{RAW}	Water flow measurement system based on a kNN classifier.
$maximum$	Maximum of the last 21 ADC_{peak} values (unit: ADC value).
$minimum$	Minimum of the last 21 ADC_{peak} values (unit: ADC value).
sum	Sum of the last 21 ADC_{peak} values (unit: ADC value).
$TestSet$	Test data set for initial system evaluation.
$thr_{powerCooling}$	Threshold used to recognize a fridge's cooling period (unit: ADC value).
U_{Ind}	Induced voltage (unit: Volt).
$variance$	Variance calculated based on the last 21 ADC_{peak} values.

Spotting and Recognition of Subtle Daily Life Arm Activities and Object Interactions

$(BS_{SA+ROI+RFL} + IS)_{subtleAct}$	The basic system $BS_{optDistOH}$ fused with operating mode information from smart appliances, sub-room level location, magnetic field signatures and a state-of-the-art inertial sensor system to identify object interactions and subtle arm actions on top of them.
$(BS_{SA+ROI+RFL} + IS + SA)_{subtleAct}$	The basic system $BS_{optDistOH}$ fused with operating mode information from smart appliances, sub-room level location, magnetic field signatures and high level operating mode information as well as a state-of-the-art inertial sensor system to identify object interactions and subtle arm actions on top of them.
$(BS_{SA+ROI+RFL} + SA)_{subtleAct}$	The basic system $BS_{optDistOH}$ fused with operating mode information from smart appliances, sub-room level location, magnetic field signatures and high level operating mode information to identify object interactions and subtle arm actions on top of them.
AI_R	Amount of analyzed images.
BS	Core components of the introduced system. The basic system includes a hand mounted camera, a proximity sensor as well as inertial sensors.
BS_{BT}	The basic system $BS_{optDistOH}$ fused with room level location information.
BS_{BT+RFL}	The basic system $BS_{optDistOH}$ fused with room level location and magnetic field signatures.
BS_{BT+ROI}	The basic system $BS_{optDistOH}$ fused with room level location and sub-room level location information.
$BS_{BT+ROI+RFL}$	The basic system $BS_{optDistOH}$ fused with room level, sub-room level and magnetic field signatures.
BS_{HM}	The basic system $BS_{optDistOH}$ fused with hand movement intensity information.

BS_{MoL}	The basic system $BS_{optDistOH}$ fused with mode of locomotion information.
$BS_{optDistHO}$	The basic system optimized in terms of $thrDistOH_{up}$ and $thrDistOH_{down}$.
BS_{RFL}	The basic system $BS_{optDistOH}$ fused with magnetic field signatures.
BS_{ROI}	The basic system $BS_{optDistOH}$ fused with sub-room level location information.
$BS_{ROI+RFL}$	The basic system $BS_{optDistOH}$ fused with sub-room level location and magnetic field signatures.
$BS_{ROI+RFL} + IS$	The basic system $BS_{optDistOH}$ fused with sub-room level location, magnetic field signatures and a state-of-the-art inertial sensor system.
BS_{SA}	The basic system $BS_{optDistOH}$ fused with operating mode information coming from smart devices.
$BS_{SA+ROI+RFL}$	The basic system $BS_{optDistOH}$ fused with operating mode information coming from smart appliances, sub-room level and magnetic field signatures.
$BS_{SA+ROI+RFL} + IS$	The basic system $BS_{optDistOH}$ fused with operating mode information from smart appliances, sub-room level location, magnetic field signatures and a state-of-the-art inertial sensor system.
BS_{TF}	The basic system $BS_{optDistOH}$ fused with time features.
$BS_{TF+MoL+RFL}$	The basic system $BS_{optDistOH}$ fused with time features, sub-room level and magnetic field signatures. This configuration does not need any infrastructure instrumentation.
$BS_{TF+MoL+RFL} + IS$	The basic system $BS_{optDistOH}$ fused with time features, sub-room level, magnetic field signatures and a state-of-the-art inertial sensor system. This configuration does not need any infrastructure instrumentation.
$BS_{TF+MoL+ROI+RFL}$	The basic system $BS_{optDistOH}$ fused with time features, modes of locomotion, sub-room level and magnetic field signatures.

$BS_{TF+MoL+SA+ROI+RFL}$	The basic system $BS_{optDistOH}$ fused with time features, modes of locomotion, operating mode information from smart appliances, sub-room level location and magnetic field signatures.
C	Ground truth event correct classified (event based evaluation)
CS_R	Amount of performed classification steps.
D	Ground truth event deleted (event based evaluation)
dr	Deletions (frame based evaluation)
EER	Overall achieved system "Equal Error Rate".
EER_{act}	Achieved EER for specific activities.
EER_{obj}	Achieved EER for specific objects.
F	Ground truth event fragmented (event based evaluation)
FM	Ground truth event fragmented and merged (event based evaluation)
FM'	Recognized event fragmented and merged (event based evaluation)
fpr	False positive rate (frame based evaluation)
fr	Fragmentations (frame based evaluation)
I'	Recognized event falsely inserted (event based evaluation)
ir	Insertions (frame based evaluation)
IS	State-of-the-art inertial sensor system using a statistically relevant amount of training data.
IS_{opt}	Inertial sensor system optimized in terms of $IS_{thrFusion}$.
$IS_{subtleAct}$	Optimized state-of-the-art inertial sensor system using a statistically relevant amount of training data - Application scenario: Subtle activity recognition.

$IS_{thrFusion}$	Time threshold used to fuse consecutive segments spotted by the inertial sensor system (ranging from 0 to 1 second, in steps of 0.2 seconds).
$IS_{thrSVMScore}$	Minimum SVM score which has to be exceeded to keep the classification result of the inertial sensor system (ranging from 0 to 1 in steps of 0.01).
$IS'_{HH_{SVMThr}}$	Minimum SVM score which has to be exceeded to keep the classification result of IS'_{HH} (ranging from 0 to 1 in steps of 0.01).
$IS'_{HH_{TS_i}}$	Interesting time sequences spotted by IS'_{HH}.
IS'_{HH}	IS_{opt} extension: Adding information about the user's hand height to reduce the amount of possible objects.
$IS'_{PROX+HH_{SVMThr}}$	Minimum SVM score which has to be exceeded to keep the classification result of $IS'_{PROX+HH}$ (ranging from 0 to 1 in steps of 0.01).
$IS'_{PROX+HH_{TS_i}}$	Interesting time sequences spotted by $IS'_{PROX+HH}$.
$IS'_{PROX+HH}$	IS_{opt} extension: Adding information about the user's hand height to reduce the amount of possible objects and object-hand distance information to spot relevant time sequences.
$List_{IS_{rankedObj}}$	Ranked list containing potential object interactions.
M	Ground truth event merged (event based evaluation)
M'	Recognized event merged (event based evaluation)
mr	Merges (frame based evaluation)
o^{α}	Start overfill (frame based evaluation)
o^{ω}	End overfill (frame based evaluation)
$precision$	Overall achieved system precision.
$precision_{act}$	Achieved precision for specific activities.
$precision_{obj}$	Achieved precision for specific objects.

$recall$	Overall achieved system recall.
$recall_{act}$	Achieved recall for specific activities.
$recall_{obj}$	Achieved recall for specific objects.
sg_x	Signal strength (in dBm) of Bluetooth beacon x.
$thr_{svmScore}$	Minimum SVM score which has to be exceeded to keep the classification result (ranging from 0 to 1 in steps of 0.01).
$thrDistOH_{down}$	Maximum allowed deviation between the average hand height of TSS_{i_j} and predefined object heights - Lower threshold (ranging from 0 cm to 40 cm in steps of 5 cm).
$thrDistOH_{up}$	Maximum allowed deviation between the average hand height of TSS_{i_j} and predefined object heights - Upper threshold (ranging from 0 cm to 40 cm in steps of 5 cm).
$thrHM_{varThr}$	Threshold on variance value of acceleration data to calculate the hand movement intensity (related to BS_{HM}; ranging from 0 m^2/s^4 to 3 m^2/s^4 in steps of 0.025 m^2/s^4).
$thrMag_{minDist}$	Threshold on closest Euclidean distance between the current average magnetic field vector and the reference vector for a specific object (related to BS_{RFL}; ranging from 0.1 mGauss to 1 mGauss in steps of 0.1 mGauss).
$thrMag_{minObjTime}$	Time threshold which must be exceeded to keep an object (related to BS_{RFL}; ranging from 0.4 seconds to 2.9 seconds in steps of 0.1 seconds).
$thrMoL_{var}$	Threshold on variance value of acceleration data used to distinguish between walking and standing activities (related to BS_{MoL}; ranging from 0.25 m^2/s^4 to 15 m^2/s^4 in steps of 0.25 m^2/s^4).
$thrROI_{maxPixDiff}$	Threshold on a pixel's difference value (in terms of difference images) used to separate foreground from background pixels (related to BS_{ROI}; ranging from 250 to 2000 in steps of 250).

$thrROI_{minActPix}$	Threshold on the amount of foreground pixels within a ROI. This threshold is used to detect ROI activities and to neglect small movements (related to BS_{ROI}; ranging from 25 to 200 in steps of 25).
$thrROI_{minTime}$	A ROI is active if more than $thrROI_{minTime}$ activities have been found within a TSS_{i_j} (related to BS_{ROI}; ranging from 1 to 6 in steps of 1).
$thrTF_{maxDur}$	Threshold on maximum time duration for each spotted TSS_{i_j} (related to BS_{TF}; ranging from 1 second to 15 seconds in steps of 1 second).
tpr	True positive rate (frame based evaluation)
TS_i	Time sequences in which the user's hand is close to an object.
TS_{IS_i}	Interesting time sequences spotted by the inertial sensor system.
TSS_{i_j}	Sub-sequence of TS_i in which the hand height deviation is less than 10cm.
u^{α}	Start underfill (frame based evaluation)
u^{ω}	End underfill (frame based evaluation)
v_{avgMag}	Average magnetic field vector (related to BS_{RFL}).

Detection of User Behavior Changes Due to Medium-Term Stress Periods.

$C1$	Behavioral feature related to the overall amount of performed phone calls.
$C2$	Behavioral feature related to the phone call contact behavior.
ROI	Region of interest.
$ROI1$	Behavioral feature related to the amount of visited ROIs.
$ROI2$	Behavioral feature related to the retention time within ROIs.
$S1$	Behavioral feature related to the overall amount of received SMS.

$S2$	Behavioral feature related to the SMS contact behavior.
$SI1$	Behavioral feature related to the overall amount of social interaction.
$SI2$	Behavioral feature related to the time spent with specific persons.

OSGi4AMI - A Standardization Approach For Services In AAL Scenarios

AAL	Ambient Assisted Living.
AppSURE	A functional service that is used to detect left-on electronic devices.
Functional Services	High-level end-user services.
OSGi	Open Services Gateway Initiative.
OSGi4AMI	A standardization approach for services and devices in OSGi related to AAL scenarios.
RG	Residential Gateway.
SOA	Software Oriented Architecture.
Technological Services	Services used to process raw sensor data and to realize basic functionalities.
ZoneSURE	A functional service that is used to monitor areas within a closed environment.

Conclusion

AAL	Ambient Assisted Living.
HMI	Human-Machine Interface.
OSGi	Open Service Gateway Initiative.
OSGi4AMI	A standardization approach for services and devices in OSGi related to AAL scenarios.
RG	Residential Gateway.
UCH	Universal Control Hub.
UI	User Interface.
URC	Universal Remote Control.

List of Figures

List of Tables

Bibliography

[AAP06] B. Alavi, N. Alsindi, and K. Pahlavan, *UWB Channel Measurements for Accurate Indoor Localization*, Military Communications Conference, 2006. MILCOM 2006. IEEE, 2006, pp. 1–7. 2.2

[AB96] B. Achermann and H. Bunke, *Combination of face classifiers for person identification*, Pattern Recognition, 1996., Proceedings of the 13th International Conference on, vol. 3, 1996, pp. 416–420 vol.3. 2.2, 2.3

[ABE+] Gregory D. Abowd, Aaron F. Bobick, Irfan A. Essa, Elizabeth D. Mynatt, and Wendy A. Rogers, *The Aware Home: A living laboratory for technologies for successful aging*, Tech. report, AAAI www.aaai.org. 1.2.2

[ABK+08] Kurt Adamer, David Bannach, Tobias Klug, Paul Lukowicz, Marco Luca Sbodio, Mimi Tresman, Andreas Zinnen, and Thomas Ziegert, *Developing a Wearable Assistant for Hospital Ward Rounds: An Experience Report*, The Internet of Things (Christian Floerkemeier, Marc Langheinrich, Elgar Fleisch, Friedemann Mattern, and Sanjay E. Sarma, eds.), Lecture Notes in Computer Science, vol. 4952, Springer Berlin Heidelberg, 2008, pp. 289–307. 1.2.1.2

[ACH+01] M. Addlesee, R. Curwen, Steve Hodges, J. Newman, P. Steggles, Andy Ward, and A. Hopper, *Implementing a sentient computing system*, Computer **34** (2001), no. 8, 50–56. 2.2

[ADSP11] Javier Guerra-Casanova Alberto De Santos, Carmen Sanchez-Avila and Gonzalo Bailador-Del Pozo, *Real-Time Stress Detection by Means of Physiological Signals*, Recent Application in Biometrics, Dr. Jucheng Yang (Ed.), 2011. 5.2, 5.3

[AGS12] Christoph Amma, Marcus Georgi, and Tanja Schultz, *Airwriting: Hands-Free Mobile Text Input by Spotting and Continuous Recognition of 3d-Space Handwriting with Inertial Sensors*, Proceedings of the 2012 16th Annual International Symposium on Wearable Computers (ISWC) (Washington, DC, USA), ISWC '12, IEEE Computer Society, 2012, pp. 52–59. 1.2.2

[AJT05] O. Amft, H. Junker, and G. Troster, *Detection of eating and drinking arm gestures using inertial body-worn sensors*, Wearable Computers, 2005. Proceedings. Ninth IEEE International Symposium on, 2005, pp. 160–163. 4.2.1, 4.3.1.1

[AKT07a] O. Amft, M. Kusserow, and G. Tröster, *Probabilistic parsing of dietary activity events*, BSN 2007: Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks (Steffen Leonhardt, Thomas Falck, and Petri Mähönen, eds.), vol. 13, IFMBE Proceedings, Springer, March 2007, pp. 242–247. 4.2.6

[AKT07b] Oliver Amft, Martin Kusserow, and Gerhard Tröster, *Automatic Identification of Temporal Sequences in Chewing Sounds.*, BIBM, IEEE Computer Society, 2007, pp. 194–201. 1.2.1.2, 4.2.6

[AMS02] Stavros Antifakos, Florian Michahelles, and Bernt Schiele, *Proactive Instructions for Furniture Assembly*, Proceedings of the 4th international conference on Ubiquitous Computing (London, UK, UK), UbiComp '02, Springer-Verlag, 2002, pp. 351–360. 1.2.1.1

[API+11] Nadav Aharony, Wei Pan, Cory Ip, Inas Khayal, and Alex Pentland, *The social fMRI: measuring, understanding, and designing social mechanisms in the real world*, Proceedings of the 13th international conference on Ubiquitous computing (New York, NY, USA), UbiComp '11, ACM, September 2011, pp. 445–454. 1.2.1.5, 5.2, 5.3

[AS02] D. Ashbrook and T. Starner, *Learning significant locations and predicting user movement with GPS*, Wearable Computers, 2002. (ISWC 2002). Proceedings. Sixth International Symposium on, 2002, pp. 101–108. 5.3

[ASLT05] Oliver Amft, Mathias Stäger, Paul Lukowicz, and Gerhard Tröster, *Analysis of chewing sounds for dietary monitoring*, Proceedings of the 7th international conference on Ubiquitous Computing (Berlin, Heidelberg), UbiComp'05, Springer-Verlag, 2005, pp. 56–72. 1.2.1.2, 1.2.2

[BABMPBMA+12] R. Bin Ambar, H. Bin Mhd Poad, A.M. Bin Mohd Ali, M.S. Bin Ahmad, and M. Mahadi bin Abdul Jamil, *Multi-sensor arm rehabilitation monitoring device*, Biomedical Engineering (ICoBE), 2012 International Conference on, 2012, pp. 424–429. 1.2.1.2

[BAF+91] Shahina Begum, Mobyen Ahmed, Peter Funk, Ning Xiong, and Bo Von Scheele, *Using Calibration and Fuzzification of Cases for Improved Diagnosis and Treatment of Stress*, INFORMATION AND COMPUTATION (1991), 93–172. 5.2

[BAK+07] David Bannach, Oliver Amft, Kai S. Kunze, Ernsta. Heinz, Gerhard Tröster, and Paul Lukowicz, *Waving Real Hand Gestures Recorded by Wearable Motion Sensors to a Virtual Car and Driver in a Mixed-Reality Parking Game*, Computational Intelligence and Games, 2007. CIG 2007. IEEE Symposium on, April 2007, pp. 32–39. 1.2.1.4

[BAL08] David Bannach, Oliver Amft, and Paul Lukowicz, *Rapid Prototyping of Activity Recognition Applications*, IEEE Pervasive Computing **7** (2008), no. 2, 22–31. 2.6.2.1, 3.4.3.2

[BAS96] Michael S Brandstein, John E Adcock, and Harvey F Silverman, *Microphone-array localization error estimation with application to sensor placement*, The Journal of the Acoustical Society of America **99** (1996), no. 6, 3807–3816. 2.2

[BAS97] M.S. Brandstein, J.E. Adcock, and H.F. Silverman, *A closed-form location estimator for use with room environment microphone arrays*, Speech and Audio Processing, IEEE Transactions on, vol. 5, Jan 1997, pp. 45–50. 2.2

[BB09] R. Belwal and S. Belwal, *Mobile Phone Usage Behavior of University Students in Oman*, New Trends in Information and Service Science, 2009. NISS '09. International Conference on, 2009, pp. 954–962. 5.3

[BBC+03] Roberto Battiti, Mauro Brunato, Renato Lo Cigno, A. Villani, R. Flor, and Gianni Lazzari, *WILMA: An Open Lab for 802.11 HotSpots: Extended abstract.*, PWC (Marco Conti, Silvia Giordano, Enrico Gregori, and Stephan Olariu, eds.), Lecture Notes in Computer Science, vol. 2775, Springer, 2003, pp. 163–168. 2.2

[BBD+11] Michelle N. Burns, Mark Begale, Jennifer Duffecy, Darren Gergle, Chris J. Karr, Emily Giangrande, and David C. Mohr, *Harnessing Context Sensing to Develop a Mobile Intervention for Depression*, Journal of Medical Internet Research, vol. 13, August 2011, pp. e55+. 1.2.1.2

[BBLS13] Gerald Bauer, Ulf Blanke, Paul Lukowicz, and Bernt Schiele, *User independent, multi-modal spotting of subtle arm actions with minimal training data*, Pervasive Computing and Communications Workshops (PERCOM Workshops), 2013 IEEE International Conference on, 2013, pp. 8–13. 1.3, 4.3.1.1

[BD13] K. Barczewska and A. Drozd, *Comparison of methods for hand gesture recognition based on Dynamic Time Warping algorithm*, Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on, 2013, pp. 207–210. 4.10, 7.2

[BFT09] Marc Bächlin, Kilian Förster, and Gerhard Tröster, *SwimMaster: A wearable assistant for swimmer*, Ubicomp '09: Proceedings of the 11th international conference on Ubiquitous computing (New York, NY, USA), ACM, 2009, pp. 215–224. 1.2.1.3

[BGGBN08] M. Berenguer, M. Giordani, F. Giraud-By, and N. Noury, *Automatic detection of activities of daily living from detecting and classifying electrical events on the residential power line*, e-health Networking, Applications and Services, 2008. HealthCom 2008. 10th International Conference on, 2008, pp. 29–32. 3.2.2

[BGL07] A. Bottaro, A. Gerodolle, and Philippe Lalanda, *Pervasive Service Composition in the Home Network*, Advanced Information Networking and Applications, 2007. AINA '07. 21st International Conference on, 2007, pp. 596–603. 6.5.3

[BHG+09] Gerald Bieber, Andre Hoffmeyer, Enrico Gutzeit, Christian Peter, and Bodo Urban, *Activity monitoring by fusion of optical and mechanical tracking technologies for user behavior analysis*, Proceedings of the 2nd International Conference on Pervasive Technologies Related to Assistive Environments (New York, NY, USA), PETRA '09, ACM, 2009, pp. 45:1–45:6. 2.2, 2.3

[BI04] Ling Bao and Stephen S. Intille, *Activity Recognition from User-Annotated Acceleration Data*, Proceedings of the 2nd International Conference on Pervasive Computing, April 2004, pp. 1–17. 4.3.1.1, 4.3.1.1

[Bin11] Max Binder, *Benutzerspezifische Mustererkennung in Telefonaten, Nachrichtenverkehr, Internetverhalten und Beschleunigungssensor-Daten bei Smartphones.*, Bachelorarbeit, Fakultät für Informatik und Mathematik, Universität Passau, 2011. 5

[Bis06] Christopher M. Bishop, *Pattern recognition and machine learning*, Springer, New York, 2006. 3.4.2.1, 3.4.2.2

[BKV+97] C.V.C. Bouten, K.T.M. Koekkoek, M. Verduin, R. Kodde, and J.D. Janssen, *A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity*, Biomedical Engineering, IEEE Transactions on **44** (1997), no. 3, 136–147. 4.7.3.1

[BKYC07] Kyu-Dae Ban, Keun-Chang Kwak, Ho-Sub Yoon, and Yun-Koo Chung, *Fusion Technique for User Identification Using Camera and Microphone in the Intelligent Service Robots*, Consumer Electronics, 2007. ISCE 2007. IEEE International Symposium on, 2007, pp. 1–6. 2.3

[BL08] Gerald Bauer and Paul Lukowicz, *Developing a Sub Room Level Indoor Location System for Wide Scale Deployment in Assisted Living Systems.*, ICCHP (Klaus Miesenberger, Joachim Klaus, Wolfgang L. Zagler, and Arthur I. Karshmer, eds.), Lecture Notes in Computer Science, vol. 5105, Springer, 2008, pp. 1057–1064. 1.1.3, 1.3, 4.7.3.1, 4.7.3.1

[BL10] Zeungnam Bien and SangWan Lee, *Learning Structure of Human Behavior Patterns in a Smart Home System*, Quantitative Logic and Soft Computing 2010 (Bing-yuan Cao, Guo-jun Wang, Shui-li Chen, and Si-zong Guo, eds.), Advances in Intelligent and Soft Computing, vol. 82, Springer Berlin Heidelberg, 2010, pp. 1–15. 1.2.2

[BL12] Gerald Bauer and Paul Lukowicz, *Can smartphones detect stress-related changes in the behaviour of individuals?*, PerCom Workshops, IEEE, 2012, pp. 423–426. 1.3

[Bla11] Ulf Mario Blanke, *Recognizing Complex Human Activity Based on Activity Spotting.*, Ph.D. thesis, Darmstadt University of Technology, 2011, http://d-nb.info/1017353484, pp. 1–174. 1.1.1, 1.2.1.1, 1.2.2, 1.2.2, 1.2, 4.1.1.1, 4.2.7

[BM98] Avrim Blum and Tom Mitchell, *Combining labeled and unlabeled data with co-training*, Morgan Kaufmann Publishers, 1998, pp. 92–100. 1.2.2, 4.1.1.1

[BOV03] A. Barla, F. Odone, and A. Verri, *Histogram intersection kernel for image classification*, Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on, vol. 3, 2003, pp. III–513–16 vol.2. 4.10, 7.2

[BP00] P. Bahl and V.N. Padmanabhan, *RADAR: An in-building RF-based user location and tracking system*, INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, vol. 2, 2000, pp. 775–784 vol.2. 2.2

[BPPW09] Michael Buettner, Richa Prasad, Matthai Philipose, and David Wetherall, *Recognizing daily activities with rfid-based sensors*, Proceedings of the 11th International Conference on Ubiquitous Computing (New York, NY, USA), Ubicomp '09, ACM, 2009, pp. 51–60. 4.2.7

[BS09] Ulf Blanke and Bernt Schiele, *Daily Routine Recognition through Activity Spotting*, 4rd International Symposium on Location- and Context-Awareness (LoCA), 2009. 4.3.1.1

[BSK$^+$10] Ulf Blanke, Bernt Schiele, Matthias Kreil, Paul Lukowicz, Bernard Sick, and Thiemo Gruber, *All for one or one for all? – Combining Heterogeneous Features for Activity Spotting*, 7th IEEE PerCom Workshop on Context Modeling and Reasoning (CoMoRea) (Mannheim, Germany), 2010. 4.3.1.1

[BSL09] Gerald Bauer, Karl Stockinger, and Paul Lukowicz, *Recognizing the Use-Mode of Kitchen Appliances from Their Current Consumption.*, EuroSSC (Payam M. Barnaghi, Klaus Moessner, Mirko Presser, and Stefan Meissner, eds.), Lecture Notes in Computer Science, vol. 5741, Springer, 2009, pp. 163–176. 1.3

[BWK08] S. Beauregard, Widyawan, and M. Klepal, *Indoor PDR performance enhancement using minimal map information and particle filters*, Position, Location and Navigation Symposium, 2008 IEEE/ION, 2008, pp. 141–147. 2.2

[CAL10] Jingyuan Cheng, Oliver Amft, and Paul Lukowicz, *Active Capacitive Sensing: Exploring a New Wearable Sensing Modality for Activity Recognition*, Pervasive, 2010, pp. 319–336. 4.2.4

[CALMT11] Burcu Cinaz, Bert Arnrich, Roberto La Marca, and Gerhard Tröster, *Monitoring of Mental Workload Levels during an Everyday Life Office-Work Scenario*, Accepted for Personal and Ubiquitous Computing Journal (2011). 5.2

[CAM+11] N. Carbonaro, G. Anania, G.D. Mura, M. Tesconi, A. Tognetti, G. Zupone, and D. De Rossi, *Wearable biomonitoring system for stress management: A preliminary study on robust ECG signal processing*, World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2011 IEEE International Symposium on a, 2011, pp. 1–6. 5.2

[CBA+08] Jingyuan Cheng, David Bannach, Kurt Adamer, Thomas Bernreiter, and Paul Lukowicz, *A Wearable, Conductive Textile Based User Interface for Hospital Ward Rounds Document Access*, Smart Sensing and Context (Daniel Roggen, Clemens Lombriser, Gerhard Tröster, Gerd Kortuem, and Paul Havinga, eds.), Lecture Notes in Computer Science, vol. 5279, Springer Berlin Heidelberg, 2008, pp. 182–191. 1.2.1.2

[CBL08] Jingyuan Cheng, D. Bannach, and P. Lukowicz, *On body capacitive sensing for a simple touchless user interface*, Medical Devices and Biosensors, 2008. ISSS-MDBS 2008. 5th International Summer School and Symposium on, 2008, pp. 113–116. 1.2.1.2

[CBL12] Jingyuan Cheng, G. Bahle, and P. Lukowicz, *A simple wristband based on capacitive sensors for recognition of complex hand motions*, Sensors, 2012 IEEE, 2012, pp. 1–4. 4.2.4, 4.10, 7.2

[CCC+09] Meng-Chieh Chiu, Shih-Ping Chang, Yu-Chen Chang, Hao-Hua Chu, Cheryl Chia-Hui Chen, Fei-Hsiu Hsiao, and Ju-Chun Ko, *Playful bottle: A mobile social persuasion system to motivate healthy water intake*, Proceedings of the 11th international conference on Ubiquitous computing (New York, NY, USA), Ubicomp '09, ACM, 2009, pp. 185–194. 1.2.1.2

[CDC10] Chao Chen, Barnan Das, and Diane J. Cook, *Energy Prediction Based on Resident's Activity*, 4th International Workshop on Knowledge Discovery from Sensor Data (SensorKDD-2010), 2010. 3.2.2

[CFJ04] Harry Chen, Tim Finin, and Anupam Joshi, *A Context Broker for Building Smart Meeting Rooms*, AAAI 2004 Spring Symposium on Knowledge Representation and Ontology for Autonomous Systems (Stanford), 2004, Draft. 1.2.1.5

[CHB+05] Trista P. Chen, Horst Haussecker, Alexander Bovyrin, Roman Belenov, Konstantin Rodyushkin, Alexander Kuranov, and Victor Eruhimov, *Computer Vision Workload Analysis: Case Study of Video Surveillance Systems*, no. 2, 109–118. 2.6.2.2

[CKZ+05] Jianfeng Chen, Alvin Harvey Kam, Jianmin Zhang, Ning Liu, and Louis Shue, *Bathroom Activity Monitoring Based on Sound*, Pervasive, 2005, pp. 47–61. 4.2.6

[CL08] Jingyuan Cheng and P. Lukowicz, *Towards wearable capacitive sensing of physiological parameters*, Pervasive Computing Technologies for Healthcare, 2008. PervasiveHealth 2008. Second International Conference on, 2008, pp. 272–273. 1.2.1.2

[CL11] C.C. Chang and C.J. Lin, *LIBSVM: a library for support vector machines*, ACM Transactions on Intelligent Systems and Technology (TIST) **2** (2011), no. 3, 27. 4.3.1.1

[CNL10] A. Czabke, J. Neuhauser, and T.C. Lueth, *Recognition of interactions with objects based on radio modules*, Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2010 4th International Conference on, March 2010, pp. 1–8. 4.2.3, 4.5.1

[CST00] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*, Cambridge University Press, 2000. 3.4.2.1

[CSV+12] Pietro Cipresso, Silvia Serino, Daniela Villani, Claudia Repetto, Luigi Sellitti, Giovanni Albani, Alessandro Mauro, Andrea Gaggioli, and Giuseppe Riva, *Is your phone so smart to affect your state? An exploratory study based on psychophysiological measures*, vol. 84, Elsevier Science Publishers B. V., May 2012, pp. 23–30. 5.2

[CSZ06] O. Chapelle, B. Schölkopf, and A. Zien (eds.), *Semi-Supervised Learning*, MIT Press, Cambridge, MA, 2006. 1.2.2, 4.1.1.1

[CT09] Min-Xiou Chen and Tze-Chin Tzeng, *Heterogeneous service location service architecture based on OSGi technology*, Advanced Communication Technology, 2009. ICACT 2009. 11th International Conference on, vol. 03, 2009, pp. 1838–1843. 6.2

[CZKS05] Jianfeng Chen, Jianmin Zhang, Alvin Harvey Kam, and Louis Shue, *An automatic acoustic bathroom monitoring system.*, ISCAS (2), IEEE, 2005, pp. 1750–1753. 3.2.1, 3.3.1

[Dal06] Navneet Dalal, *Finding people in images and videos*, Ph.D. thesis, Institut National Polytechnique de Grenoble, July 2006. 1.2.2, 4.1.1.1, 4.2.2, 4.5.1.3, 4.7.1.2, 4.7.1.2

[DBL11] B. Sick D. Bannach and P. Lukowicz, *Automatic Adaptation of Mobile Activity Recognition Systems to New Sensors*, Workshop Mobile Sensing: Challenges, Opportunities, and Future Directions (UbiComp 11), ACM, 2011. 1.2.2, 4.1.1.1

[DD01] R.W. DeVaul and S. Dunn, *Real-time motion classification for wearable computing applications*, 2001, project paper, http://www. media. mit. edu/wearables/mithril/realtime. pdf (2001). 4.3.1.1

[DDS+09] Jia Deng, Wei Dong, R. Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, *ImageNet: A large-scale hierarchical image database*, Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, 2009, pp. 248–255. 4.2.2

[DSAF99] A.K. Dey, D. Salber, G.D. Abowd, and M. Futakawa, *The Conference Assistant: Combining context-awareness with wearable computing*, Wearable Computers, 1999. Digest of Papers. The Third International Symposium on, 1999, pp. 21–28. 1.2.1.5

[dSSAGC+10] A. de Santos Sierra, C.S. Avila, J. Guerra Casanova, G.B. del Pozo, and V.J. Vera, *Two Stress Detection Schemes Based on Physiological Signals for Real-Time Applications*, Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010 Sixth International Conference on, 2010, pp. 364–367. 5.2

[dSSSACdP11] A. de Santos Sierra, C. Sanchez Avila, J.G. Casanova, and G.B. del Pozo, *A Stress-Detection System Based on Physiological Signals and Fuzzy Logic*, Industrial Electronics, IEEE Transactions on **58** (2011), no. 10, 4857 –4865. 5.2, 5.3

[DT05] N. Dalal and B. Triggs, *Histograms of oriented gradients for human detection*, Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 1, 2005, pp. 886–893 vol. 1. 1.2.2, 4.2.2, 4.5.1.3, 4.5.5, 4.7.1.1, 4.7.1.2, 4.7.1.2

[EPL09] Nathan Eagle, Alex S. Pentland, and David Lazer, *From the Cover: Inferring friendship network structure by using mobile phone data*, Proceedings of the National Academy of Sciences **106** (2009), no. 36, 15274–15278. 1.2.1.5, 5.2, 5.3

[EZW+06] Mark Everingham, Andrew Zisserman, Christopher K. I. Williams, Luc Van Gool, Moray Allan, Christopher M. Bishop, Olivier Chapelle, Navneet Dalal, Thomas Deselaers, Gyuri Dorkó, Stefan Duffner, Jan Eichhorn, Jason D. R. Farquhar, Mario Fritz, Christophe Garcia, Tom Griffiths, Frédéric Jurie, Thomas Keysers, Markus Koskela, Jorma Laaksonen, Diane Larlus, Bastian Leibe, Hongying Meng, Hermann Ney, Bernt Schiele, Cordelia Schmid, Edgar Seemann, John Shawe-Taylor, Amos Storkey, Sandor Szedmak, Bill Triggs, Ilkay Ulusoy, Ville Viitaniemi, and Jianguo Zhang, *The 2005 PASCAL Visual Object Classes Challenge*, LNAI, Springer, 2006. 4.2.2, 4.5.1.3

[FAH06] James Fogarty, Carolyn Au, and Scott E. Hudson, *Sensing from the basement: A feasibility study of unobtrusive and low-cost home activity recognition.*, UIST (Pierre Wellner and Ken Hinckley, eds.), ACM, 2006, pp. 91–100. 3.2.1, 3.3, 3.3.1

[FFFP03] Li Fei-Fei, Rob Fergus, and Pietro Perona, *A Bayesian Approach to Unsupervised One-Shot Learning of Object Categories*, Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2 (Washington, DC, USA), ICCV '03, IEEE Computer Society, 2003, pp. 1134–. 1.2.2, 1.2, 4.1.1.1

[FFFP06] Li Fei-Fei, R. Fergus, and P. Perona, *One-shot learning of object categories*, Pattern Analysis and Machine Intelligence, IEEE Transactions on **28** (2006), no. 4, 594–611. 1.2.2, 4.1.1.1

[Fin11] Alexander Findeis, *Benutzerspezifische Mustererkennung in GPS-, WLAN- und Bluetoothdaten von Smartphones.*, Bachelorarbeit, Fakultät für Informatik und Mathematik, Universität Passau, 2011. 5, 5.5.1, 5.2, 60, 7.3.3

[FKW+10] Gunnar Fagerberg, Antonio Kung, Reiner Wichert, Mohammad-Reza Tazari, Bruno Jean-Bart, Gerald Bauer, Gottfried Zimmermann, Francesco Furfari, Francesco Potorti, Stefano Chessa, Michael Hellenschmidt, Joe Gorman, Jan Alexandersson, Jürgen Bund, Eduardo Carrasco, Gorka Epelde, Martin Klima, Elena Urdaneta, Gregg C. Vanderheiden, and Ingo Zinnikus, *Platforms for AAL Applications.*, EuroSSC (Paul Lukowicz, Kai S. Kunze, and Gerd Kortuem, eds.), Lecture Notes in Computer Science, vol. 6446, Springer, 2010, pp. 177–201. 1.3, 7.3

[FLKB09] T. Franke, P. Lukowicz, K. Kunze, and D. Bannach, *Can a Mobile Phone in a Pocket Reliably Recognize Ambient Sounds?*, Wearable Computers, 2009. ISWC '09. International Symposium on, 2009, pp. 161–162. 4.2.6

[FMH+10] T. Fuxreiter, C. Mayer, S. Hanke, M. Gira, M. Sili, and J. Kropf, *A modular platform for event recognition in smart homes*, e-Health Networking Applications and Services (Healthcom), 2010 12th IEEE International Conference on, 2010, pp. 1–6. 6.2

[FMT+99] J. Farringdon, A.J. Moore, N. Tilbury, J. Church, and P.D. Biemond, *Wearable sensor badge and sensor jacket for context awareness*, Wearable Computers, 1999. Digest of Papers. The Third International Symposium on, 1999, pp. 107–113. 1.2.1.3, 2.3, 4.2.1

[Fox05] E. Foxlin, *Pedestrian tracking with shoe-mounted inertial sensors*, Computer Graphics and Applications, IEEE **25** (2005), no. 6, 38–46. 2.2

[FZ09] A. Ferscha and K. Zia, *LifeBelt: Silent Directional Guidance for Crowd Evacuation*, Wearable Computers, 2009. ISWC '09. International Symposium on, 2009, pp. 19–26. 1.2.1.6

[GOB+12] Agnes Grünerbl, Patricia Oleksy, Gernot Bahle, Christian Haring, Jens Weppner, and Paul Lukowicz, *Towards smart phone based monitoring of bipolar disorder*, Proceedings of the Second ACM Workshop on Mobile Systems, Applications, and Services for HealthCare (New York, NY, USA), mHealthSys '12, ACM, 2012, pp. 3:1–3:6. 1.2.1.2

[GPBB+13] Tobias Grosse-Puppendahl, Yannick Berghoefer, Andreas Braun, Raphael Wimmer, and Arjan Kuijper, *OpenCapSense: A Rapid Prototyping Toolkit for Pervasive Interaction using Capacitive Sensing*, Pervasive Computing and Communications Workshops (PERCOM Workshops), 2013 IEEE International Conference on, 2013, pp. 151–158. 2.2

[GPT+] Andrea Gaggioli, Giovanni Pioggia, Gennaro Tartarisco, Giovanni Baldus, Daniele Corda, Pietro Cipresso, and Giuseppe Riva, *A mobile data collection platform for mental health research*, Personal and Ubiquitous Computing, 1–11, 10.1007/s00779-011-0465-2. 5.2, 5.8, 7.2

[GVG07] Charles Gouin-Vallerand and Sylvain Giroux, *Managing and Deployment of Applications with OSGi in the Context of Smart Home.*, WiMob, IEEE Computer Society, 2007, p. 70. 6.5

[GWM11] Dominic Gorecky, Simon F. Worgan, and Gerrit Meixner, *COGNITO: A cognitive assistance and training system for manual tasks in industry*, Proceedings of the 29th Annual European Conference on Cognitive Ergonomics (New York, NY, USA), ECCE '11, ACM, 2011, pp. 53–56. 4.5.7

[GYL+07] Donghai Guan, Weiwei Yuan, Young-Koo Lee, A. Gavrilov, and Sungyoung Lee, *Activity Recognition Based on Semi-supervised Learning*, Embedded and Real-Time Computing Systems and Applications, 2007. RTCSA 2007. 13th IEEE International Conference on, 2007, pp. 469–475. 1.2.2, 4.1.1.1

[HB01] Jeffrey Hightower and G. Borriello, *Location systems for ubiquitous computing*, Computer **34** (2001), no. 8, 57–66. 2.2

[HC10] Chen Hang and Cao Can, *Research and Application of Distributed OSGi for Cloud Computing*, Computational Intelligence and Software Engineering (CiSE), 2010 International Conference on, 2010, pp. 1–5. 6.2

[HCI+12] S. Hinterstoisser, C. Cagniart, S. Ilic, P. Sturm, N. Navab, P. Fua, and V. Lepetit, *Gradient Response Maps for Real-Time Detection of Textureless Objects*, Pattern Analysis and Machine Intelligence, IEEE Transactions on **34** (2012), no. 5, 876–888. 4.2.2, 4.5.1.3

[HCL+05] Jeffrey Hightower, Sunny Consolvo, Anthony LaMarca, Ian Smith, and Jeff Hughes, *Learning and recognizing the places we go*, Proceedings of the 7th international conference on Ubiquitous Computing (Berlin, Heidelberg), UbiComp'05, Springer-Verlag, 2005, pp. 159–176. 5.3

[HFS08] Tâm Huynh, Mario Fritz, and Bernt Schiele, *Discovery of Activity Patterns using Topic Models*, Proceedings of the 10th ACM International Conference on Ubiquitous Computing (UbiComp), 2008. 4.3.1.1

[HHC+11] Stefan Hinterstoisser, Stefan Holzer, Cedric Cagniart, Slobodan Ilic, Kurt Konolige, Nassir Navab, and Vincent Lepetit, *Multimodal templates for real-time detection of texture-less objects in heavily cluttered scenes*, ICCV, 2011, pp. 858–865. 4.2.2, 4.5.1.3

[HKG+06] E.A. Heinz, K.S. Kunze, M. Gruber, D. Bannach, and P. Lukowicz, *Using Wearable Sensors for Real-Time Recognition Tasks in Games of Martial Arts - An Initial Experiment*, Computational Intelligence and Games, 2006 IEEE Symposium on, 2006, pp. 98–102. 1.2.1.3, 4.2.1

[HLI+10] Stefan Hinterstoisser, Vincent Lepetit, Slobodan Ilic, Pascal Fua, and Nassir Navab, *Dominant orientation templates for real-time detection of texture-less objects*, CVPR, 2010, pp. 2257–2264. 4.2.2

[HLMA09] Ling He, M. Lech, N.C. Maddage, and N. Allen, *Stress Detection Using Speech Spectrograms and Sigma-pi Neuron Units*, Natural Computation, 2009. ICNC '09. Fifth International Conference on, vol. 2, aug. 2009, pp. 260 –264. 5.2, 5.3, 5.8

[HLS+12] Kuo-Hsun Hsu, Wen-Tin Lee, Min-Yu Sie, Guan-Lin Ciou, and Shao-Yuan Lu, *Extending OSGi with Instant Messaging Communications and Peer to Peer Transfer Based on XMPP*, Computer, Consumer and Control (IS3C), 2012 International Symposium on, 2012, pp. 650–653. 6.2

[HS05] T. Huynh and B. Schiele, *Analyzing Features for Activity Recognition*, Proceedings of the ACM International Conference of the joint conference on Smart objects and ambient intelligence (EUSAI) (Grenoble, France), 2005. 4.3.1.1

[HVBW01] J. Hightower, C. Vakili, G. Borriello, and R. Want, *Design and Calibration of the SpotON Ad-Hoc Location Sensing System*, 2001. 2.2

[HWB00] J. Hightower, R. Want, and G. Borriello, *SpotON: An indoor 3D location sensing technology based on RF signal strength*, UW CSE 00-02-02, (2000). 2.2

[HWL+03] S. Helal, B. Winkler, Choonhwa Lee, Y. Kaddoura, L. Ran, C. Giraldo, S. Kuchibhotla, and W. Mann, *Enabling location-aware pervasive computing applications for the elderly*, Pervasive Computing and Communications, 2003. (PerCom 2003). Proceedings of the First IEEE International Conference on, 2003, pp. 531–536. 6.2

[HWT11] Thomas Holleczek, Martin Wirz, and Gerhard Tröster, *Towards Collision Avoidance on Ski Slopes*, Proceedings of the 19th International Congress on Ski Trauma and Skiing Safety (ISSS 2011), May 2011. 1.2.1.3

[HYI04] T. Hirayama, M. Yachida, and Y. Iwai, *Parallelization between face localization and person identification*, Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on, 2004, pp. 183–188. 2.2, 2.3

[Iba02] Miriam Ibanez, *Service Provisioning for the residential environment.*, OSGi Alliance Congress (2002). 6.2, 6.6

[IBC+08] Alejandro Ibarz, Gerald Bauer, Roberto Casas, Alvaro Marco, and Paul Lukowicz, *Design and Evaluation of a Sound Based Water Flow Measurement System.*, EuroSSC (Daniel Roggen, Clemens Lombriser, Gerhard Tröster, Gerd Kortuem, and Paul J. M. Havinga, eds.), Lecture Notes in Computer Science, vol. 5279, Springer, 2008, pp. 41–54. 1.3

[IIN07] T. Ikeda, H. Ishiguro, and T. Nishimura, *People tracking by cross modal association of vision sensors and acceleration sensors*, Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on, 2007, pp. 4147–4151. 2.2, 2.3

[ILB+05] Stephen S. Intille, Kent Larson, J. S. Beaudin, J. Nawyn, E. Munguia Tapia, and P. Kaushik, *A living laboratory for the design and evaluation of ubiquitous computing technologies*, CHI '05 Extended Abstracts on Human Factors in Computing Systems (New York, NY, USA), CHI EA '05, ACM, 2005, pp. 1941–1944. 1.2.2

[ILT+06] Stephen S. Intille, Kent Larson, Emmanuel Munguia Tapia, Jennifer S. Beaudin, Pallavi Kaushik, Jason Nawyn, and Randy Rockinson, *Using a live-in laboratory for ubiquitous computing research*, Proceedings of the 4th international conference on Pervasive Computing (Berlin, Heidelberg), PERVASIVE'06, Springer-Verlag, 2006, pp. 349–365. 1.2.2, 1.4

[IMMR10] Y. Ishiguro, A. Mujibiya, T. Miyaki, and J. Rekimoto, *Aided eyes: Eye activity sensing for daily life*, Proceedings of the 1st Augmented Human International Conference, ACM, 2010, p. 25. 4.2.2

[JCH+04] Xiaodong Jiang, Nicholas Y. Chen, Jason I. Hong, Kevin Wang, Leila Takayama, and James A. Landay, *Siren: Context-aware Computing for Firefighting*, in proceedings of second international conference on pervasive computing (PERVASIVE 2004), VOLUME 3001 of lecture notes in computer science, Springer, 2004, pp. 87–105. 1.2.1.6

[JCL] Kai Kunze Carl Christian Rheinländer Sebastian Wille Norbert Wehn Jens Weppner Jingyuan Cheng, Bo Zhou and Paul Lukowicz, *Activity Recognition and Nutrition Monitoring in Every Day Situations with a Textile Capacitive Neckband*, to appear at Ubicomp 2013. 1.2.1.2

[JGW04] A. Bharucha J. Gao, A. Hauptmann and H. Wactlar, *Dining activity analysis using a hidden markov model*, in 17th International Conference on Pattern Recognition (ICPR), 2004. 1.2.1.2

[JKC07] Do-Un Jeong, Se-Jin Kim, and Wan-Young Chung, *Classification of Posture and Movement Using a 3-axis Accelerometer*, Proceedings of the 2007 International Conference on Convergence Information Technology (Washington, DC, USA), ICCIT '07, IEEE Computer Society, 2007, pp. 837–844. 2.3, 4.2.1

[JLP06] Guang Y. Jin, Xiao Y. Lu, and Myong S. Park, *An Indoor Localization Mechanism Using Active RFID Tag*, Sensor Networks, Ubiquitous, and Trustworthy Computing, International Conference on **1** (2006), 40–43. 2.2

[JP11] Weppner Jens and Lukowicz Paul, *Collaborative Crowd Density Estimation with Mobile Phones*, PhoneSense '11, ACM, 2011. 1.1.3, 1.2.1.6, 5.3, 5.5.2

[JPS+13] Weppner Jens, Lukowicz Paul, Serino Silvia, Cipresso Pietro, Gaggioli Andrea, and Riva Giuseppe, *Smartphone Based Experience Sampling of Stress-Related Events*, PervasiveHealth, MindCare Workshop, 2013. 1.2.1.2, 5.2, 5.8, 7.2

[JRAJL04] Falcó Jorge, Casas Roberto, Marco Álvaro, and Sevillano José Luis, *Puede La Domótica Ayudar A Personas Con Discapacidad?*, Dómotica En Entornos Asistenciales (2004), 102–106. 6.2, 6.6

[JZZC12] Yang Ji, Chunhong Zhang, Zhihao Zuo, and Jing Chang, *Mining user daily behavior based on location history*, Communication Technology (ICCT), 2012 IEEE 14th International Conference on, 2012, pp. 881–886. 5.3

[KBH+06] K. Kunze, M. Barry, E.A. Heinz, P. Lukowicz, D. Majoe, and J. Gutknecht, *Towards Recognizing Tai Chi - An Initial Experiment Using Wearable Sensors*, Applied Wearable Computing (IFAWC), 2006 3rd International Forum on, 2006, pp. 1–6. 1.2.1.3, 1.2.2, 4.2.1, 4.3

[KHM+00] J. Krumm, S. Harris, B. Meyers, B. Brumitt, M. Hale, and S. Shafer, *Multi-camera multi-person tracking for EasyLiving*, Visual Surveillance, 2000. Proceedings. Third IEEE International Workshop on, 2000, pp. 3–10. 2.2

[KHP07] Gunhee Kim, Martial Hebert, and Sung-Kee Park, *Preliminary Development of a Line Feature-Based Object Recognition System for Textureless Indoor Objects*, Recent Progress in Robotics: Viable Robotic Service to Human (2007). 4.2.2, 4.5.1.3

[Kla07] Markus Klann, *Playing with Fire: User-Centered Design of Wearable Computing for Emergency Response*, Mobile Response, First International Workshop on Mobile Information Technology for Emergency Response, Mobile Response 2007, Sankt Augustin, Germany, February 22-23, 2007, Revised Selected Papers (Jobst Löffler and Markus Klann, eds.), Lecture Notes in Computer Science, vol. 4458, Springer, 2007, pp. 116–125. 1.2.1.6

[Kla08] Markus Klann, *Tactical Navigation Support for Firefighters: The LifeNet Ad-Hoc Sensor-Network and Wearable System.*, Mobile Response (Jobst Löffler and Markus Klann, eds.), Lecture Notes in Computer Science, vol. 5424, Springer, 2008, pp. 41–56. 1.2.1.6

[Kle08] Rolf Klein, *Algorithmische Geometrie*, vol. 2, Springer, 2008. 2.3, 2.5.3

[KLF11] K. Kloch, P. Lukowicz, and C. Fischer, *Collaborative PDR Localisation with Mobile Phones*, Wearable Computers (ISWC), 2011 15th Annual International Symposium on, 2011, pp. 37–40. 2.2

[KMSW05] Florian Kraft, Robert Malkin, Thomas Schaaf, and Alex Waibel, *Temporal ICA for Classification of Acoustic Events in a Kitchen Environment*, in Proceedings of the INTERSPEECH, 2005. 3.2.1, 3.3, 3.3.1

[KNM+06] D.M. Karantonis, M.R. Narayanan, M. Mathie, N.H. Lovell, and B.G. Celler, *Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring*, Information Technology in Biomedicine, IEEE Transactions on **10** (2006), no. 1, 156–167. 4.7.3.1

[KOA+99] Cory D. Kidd, Robert Orr, Gregory D. Abowd, Christopher G. Atkeson, Irfan A. Essa, Blair MacIntyre, Elizabeth D. Mynatt, Thad Starner, and Wendy Newstetter, *The Aware Home: A Living Laboratory for Ubiquitous Computing Research*, Proceedings of the Second International Workshop on Cooperative Buildings, Integrating Information, Organization, and Architecture (London, UK, UK), CoBuild '99, Springer-Verlag, 1999, pp. 191–198. 1.2.2

[KOL08] M. Kreil, G. Ogris, and P. Lukowicz, *Muscle activity evaluation using force sensitive resistors*, Medical Devices and Biosensors, 2008. ISSS-MDBS 2008. 5th International Summer School and Symposium on, 2008, pp. 107–110. 1.2.1.3

[KR08] B. Krach and P. Robertson, *Integration of foot-mounted inertial sensors into a Bayesian location estimation framework*, Positioning, Navigation and Communication, 2008. WPNC 2008. 5th Workshop on, 2008, pp. 55–61. 2.2

[KRG+] Markus Klann, Till Riedel, Hans Gellersen, Carl Fischer, Gerald Pirkl, Kai Kunze, Monty Beuster, Michael Beigl, Otto Visser, and Mirco Gerling, *LifeNet: An Ad-hoc Sensor Network and Wearable System to Provide Firefighters with Navigation Support.* 1.2.1.6

[KSJ10] Antonio Krüger, Lübomira Spassova, and Ralf Jung, *Innovative Retail Laboratory - Investigating Future Shopping Technologies*, it - Information Technology **52** (2010), no. 2, 114–119. 1.2.2

[Kun11] Kai Kunze, *Compensating for On-Body Placement Effects in Activity Recognition*, Ph.D. thesis, Universität Passau, Innstrasse 29, 94032 Passau, 2011. 1.1.2, 1.2.2, 4.1.1.3

[KVLL13] Matthias Kreil, Kristof Van Laerhoven, and Paul Lukowicz, *Allowing early inspection of activity data from a highly distributed bodynet with a hierarchical-clustering-of-segments approach*, Body Sensor Networks (BSN), 2013 IEEE International Conference on, 2013, pp. 1–6. 7.2

[KWSB04] Jong Hee Kang, William Welbourne, Benjamin Stewart, and Gaetano Borriello, *Extracting places from traces of locations*, Proceedings of the 2nd ACM international workshop on Wireless mobile applications and services on WLAN hotspots (New York, NY, USA), WMASH '04, ACM, 2004, pp. 110–118. 5.3

[LARC10] Paul Lukowicz, Oliver Amft, Daniel Roggen, and Jingyuan Cheng, *On-Body Sensing: From Gesture-Based Input to Activity-Driven Interaction*, IEEE Computer **43** (2010), no. 10, 92–96. 4.2.4

[LBG+09] M. Lapinski, E. Berkson, T. Gill, M. Reinold, and J.A. Paradiso, *A Distributed Wearable, Wireless Sensor System for Evaluating Professional Baseball Pitchers and Batters*, Wearable Computers, 2009. ISWC '09. International Symposium on, 2009, pp. 131–138. 1.2.1.3

[LCK+05] Jonathan Lester, Tanzeem Choudhury, Nicky Kern, Gaetano Borriello, and Blake Hannaford, *A hybrid discriminative/generative approach for modeling human activities*, Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI), 2005. 4.3.1.1

[LFO+07] Joshua Lifton, Mark Feldmeier, Yasuhiro Ono, Cameron Lewis, and Joseph A. Paradiso, *A platform for ubiquitous sensor deployment in occupational and domestic environments*, Proceedings of the 6th international conference on Information processing in sensor networks (New York, NY, USA), IPSN '07, ACM, 2007, pp. 119–127. 3.2.2, 3.3

[LHP+07] Beth Logan, Jennifer Healey, Matthai Philipose, Emmanuel Munguia Tapia, and Stephen S. Intille, *A Long-Term Evaluation of Sensing Modalities for Activity Recognition.*, Ubicomp (John Krumm, Gregory D. Abowd, Aruna Seneviratne, and Thomas Strang, eds.), Lecture Notes in Computer Science, vol. 4717, Springer, 2007, pp. 483–500. 1.1.3, 1.2.2, 1.2, 3.2.2, 3.3

[LLZ+11] Yuanqing Lin, Fengjun Lv, Shenghuo Zhu, Ming Yang, Timothee Cour, Kai Yu, Liangliang Cao, and Thomas S. Huang, *Large-scale image classification: Fast feature extraction and SVM training*, CVPR'11, 2011, pp. 1689–1696. 4.2.2, 4.5.1.3

[LM98a] H. Liu and H. Motada, *Feature Selection for Knowledge Discovery and Data Mining*, Kluwer Academic Publishers, 1998. 3.4.2.1

[LM98b] Huan Liu and Hiroshi Motoda, *Feature extraction, construction and selection: A data mining perspective*, vol. SECS 453, Kluwer Academic, Boston, 1998, edited by Huan Liu and Hiroshi Motoda.; Includes bibliographical references and index. 3.4.2.1

[LM02] Seon-Woo Lee and K. Mase, *Activity and location recognition using wearable sensors*, Pervasive Computing, IEEE **1** (2002), no. 3, 24–32. 1.2.1.3, 2.3, 2.5.2, 4.2.1, 4.5.3.2

[LNH03] Choonhwa Lee, D. Nordstedt, and S. Helal, *Enabling smart spaces with OSGi*, Pervasive Computing, IEEE **2** (2003), no. 3, 89–94. 6.5.3

[LNV+06] G. LeBellego, N. Noury, G. Virone, M. Mousseau, and J. Demongeot, *A model for the measurement of patient activity in a hospital suite*, Information Technology in Biomedicine, IEEE Transactions on **10** (2006), no. 1, 92–99. 4.2.5

[LNZ12] Qiang Lin, Hongbo Ni, and Xingshe Zhou, *An OSGi-based heath service platform for elderly people*, e-Health Networking, Applications and Services (Healthcom), 2012 IEEE 14th International Conference on, 2012, pp. 317–320. 6.2

[Low99] D.G. Lowe, *Object recognition from local scale-invariant features*, Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, vol. 2, 1999, pp. 1150–1157 vol.2. 4.2.2, 4.5.1.3, 4.5.5

[LPB+10] Paul Lukowicz, Gerald Pirkl, David Bannach, Florian Wagner, Alberto Calatroni, Kilian Förster, Thomas Holleczek, Mirco Rossi, Daniel Roggen, Gerhard Tröster, Jakob Doppler, Clemens Holzmann, Andreas Riener, Alois Ferscha, and Ricardo Chavarriaga, *Recording a Complex, Multi Modal Activity Data Set for Context Recognition*, ARCS Workshops, 2010, pp. 161–166. 1.1.2, 1.1.3, 1.2.2, 1.2, 4.5, 4.5.1.2

[LSB08] Marcus Liwicki, Andreas Schlapbach, and Horst Bunke, *Writer-Dependent Recognition of Handwritten Whiteboard Notes in Smart Meeting Room Environments*, Proceedings of the 2008 The Eighth IAPR International Workshop on Document Analysis Systems (Washington, DC, USA), DAS '08, IEEE Computer Society, 2008, pp. 151–157. 1.2.1.2

[LST13] C. Lauterbach, A. Steinhage, and A. Techmer, *A Large-Area Sensor System Underneath the Floor for Ambient Assisted Living Applications*, Pervasive and Mobile Sensing and Computing for Healthcare (Subhas Chandra Mukhopadhyay and Octavian A. Postolache, eds.), Smart Sensors, Measurement and Instrumentation, vol. 2, Springer Berlin Heidelberg, 2013, pp. 69–87. 4.7.3.1

[LWJ+04] Paul Lukowicz, Jamie A. Ward, Holger Junker, Mathias Stäger, Gerhard Tröster, Amin Atrash, and Thad Starner, *Recognizing workshop activity using body worn microphones and accelerometers*, In Pervasive Computing, 2004, pp. 18–32. 1.2.1.1

[LX96] C. Lee and Y. Xu, *Online, interactive learning of gestures for human/robot interfaces*, IEEE International Conference on Robotics and Automation, Citeseer, 1996, pp. 2982–2987. 4.3.1.1

[MBC09] A. Morrison, M. Bell, and M. Chalmers, *Visualisation of Spectator Activity at Stadium Events*, Information Visualisation, 2009 13th International Conference, 2009, pp. 219–226. 1.2.1.6

[MBL11] Michael Muehlbauer, Gernot Bahle, and Paul Lukowicz, *What Can an Arm Holster Worn Smart Phone Do for Activity Recognition?*, ISWC, 2011, pp. 79–82. 1.2.1.3, 1.2.2

[MBP14] Kreil Matthias, Sick Bernard, and Lukowicz Paul, *Dealing with human variability in motion based, wearable activity recognition*, 1th Symposium on Activity and Context Modeling and Recognition (ACOMORE) (Budapest, Hungary), 2014. 4.3

[MCB+09] Alvaro Marco, Roberto Casas, Gerald Bauer, Ruben Blasco Marin, Angel Asensio, Bruno Jean-Bart, and Miriam Ibanez, *Common OSGi Interface for Ambient Assisted Living Scenarios.*, BMI Book (Björn Gottfried and Hamid K. Aghajan, eds.), Ambient Intelligence and Smart Environments, vol. 3, IOS Press, 2009, pp. 336–357. 1.3, 72, 6.2, 6.7, 73, 6.4, 7.3.3

[MCLC04] M.J. Mathie, B.G. Celler, N.H. Lovell, and A.C.F. Coster, *Classification of basic daily movements using a triaxial accelerometer*, Medical and Biological Engineering and Computing **42** (2004), no. 5, 679–687 (English). 4.7.3.1

[MF09] T. Mitsui and H. Fujiyoshi, *Object detection by joint features based on two-stage boosting*, Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on, 2009, pp. 1169–1176. 4.10, 7.2

[MHS01] J. Mantyjarvi, J. Himberg, and T. Seppanen, *Recognizing human motion with multiple acceleration sensors*, Systems, Man, and Cybernetics, 2001 IEEE International Conference on, vol. 2, 2001, pp. 747–752 vol.2. 1.2.1.3, 2.3, 2.5.2, 4.2.1, 4.3, 4.5.3.2

[MJC+02] M. Milenkovic, E. Jovanov, J. Chapman, D. Raskovic, and J. Price, *An accelerometer-based physical rehabilitation system*, System Theory, 2002. Proceedings of the Thirty-Fourth Southeastern Symposium on, 2002, pp. 57–60. 1.2.1.2

[MKYS12] T. Maekawa, Y. Kishino, Y. Yanagisawa, and Y. Sakurai, *WristSense: Wrist-worn sensor device with camera for daily activity recognition*, Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on, 2012, pp. 510–512. 4.2.7

[MMPS10] T. Massey, G. Marfia, M. Potkonjak, and M. Sarrafzadeh, *Experimental analysis of a mobile health system for mood disorders*, Information Technology in Biomedicine, IEEE Transactions on **14** (2010), no. 2, 241–247. 1.2.1.2

[MPK10] Gerrit Meixner, Nils Petersen, and Holger Koessling, *User interaction evolution in the Smart Factory KL*, Proceedings of the 24th BCS Interaction Specialist Group Conference (Swinton, UK, UK), BCS '10, British Computer Society, 2010, pp. 211–220. 1.2.2

[MPP01] Anuj Mohan, Constantine Papageorgiou, and Tomaso Poggio, *Example-Based Object Detection in Images by Components*, IEEE Trans. Pattern Anal. Mach. Intell. **23** (2001), no. 4, 349–361. 1.2.2

[MVMJMC03] Ibanez Miram, Garcia Victor Manuel, Montero Jose Maria, and Diaz Cristina, *Aplicacion del est andar OSGi en la arquitectura del proyecto Hogar.es*, Comunicaciones de Telefonica I+D, 31 (2003), 25–34. 6.2

[MYK+10] Takuya Maekawa, Yutaka Yanagisawa, Yasue Kishino, Katsuhiko Ishiguro, Koji Kamei, Yasushi Sakurai, and Takeshi Okadome, *Object-Based Activity Recognition with Heterogeneous Sensors on Wrist.*, Pervasive (Patrik Floreen, Antonio Krüger, and Mirjana Spasojevic, eds.), Lecture Notes in Computer Science, vol. 6030, Springer, 2010, pp. 246–264. 4.2.7, 4.5.1, 4.5.1.3

[NCLZ11] Qiong Ning, Yiqiang Chen, Junfa Liu, and Huiguo Zhang, *Heterogeneous multimodal sensors based activity recognition system*, Multimedia and Expo (ICME), 2011 IEEE International Conference on, 2011, pp. 1–4. 4.2.7, 4.5.1

[NDA06] C. Nerguizian, C. Despins, and S. Affes, *Geolocation in mines with an impulse response fingerprinting technique and neural networks*, Wireless Communications, IEEE Transactions on **5** (2006), no. 3, 603–611. 2.2

[NDRCT12] Long-Van Nguyen-Dinh, D. Roggen, A. Calatroni, and G. Troster, *Improving online gesture recognition with template matching methods in accelerometer data*, Intelligent Systems Design and Applications (ISDA), 2012 12th International Conference on, 2012, pp. 831–836. 4.10, 7.2

[NK07a] T. Nicolai and H. Kenn, *About the relationship between people and discoverable Bluetooth devices in urban environments*, Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology, ACM, 2007, pp. 72–78. 5.5.2

[NK07b] Tom Nicolai and Holger Kenn, *About the relationship between people and discoverable Bluetooth devices in urban environments*, Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology (New York, NY, USA), Mobility '07, ACM, 2007, pp. 72–78. 1.2.1.6

[NLLP04] Lionel M. Ni, Yunhao Liu, Yiu C. Lau, and Abhishek P. Patil, *LANDMARC: Indoor Location Sensing Using Active RFID*, Wireless Networks **10** (2004), no. 6, 701–710. 2.2

[NSW+06] Tom Nicolai, Thomas Sindt, Hendrik Witt, Jörn Reimerdes, and Holger Kenn, *Wearable computing for aircraft maintenance: Simplifying the user interface*, In: Proceedings of the 3rd International Forum on Applied Wearable Computing (IFAWC, Verlag, 2006, pp. 15–16. 1.2.1.1

[OA00] Robert J. Orr and Gregory D. Abowd, *The smart floor: a mechanism for natural user identification and tracking*, CHI '00 Extended Abstracts on Human Factors in Computing Systems (New York, NY, USA), CHI EA '00, ACM, 2000, pp. 275–276. 2.2

[Ogr09] Georg Ogris, *Multi-modal on-body sensing of human activities*, Ph.D. thesis, 2009. 1.1.2, 1.2.1.1, 1.2.1.5, 1.2.2, 1.2, 4.1.1.1

[OSG07] OSGi Alliance, http://www.osgi.org, *OSGi Service Platform - Core Specification, Release 4.*, 2007. 6.1, 6.5, 6.5.1, 6.5.3

[OSLT08] G. Ogris, T. Stiefmeier, P. Lukowicz, and G. Tröster, *Using a complex multi-modal on-body sensor system for activity spotting*, Wearable Computers, 2008. ISWC 2008. 12th IEEE International Symposium on, 2008, pp. 55–62. 4.1.1.3, 4.2.7, 4.3, 4.2

[Par03] R. Paradiso, *Wearable health care system for vital signs monitoring*, Information Technology Applications in Biomedicine, 2003. 4th International IEEE EMBS Special Topic Conference on, 2003, pp. 283–286. 1.2.1.2

[PBLS10] R. Paradiso, A.M. Bianchi, K. Lau, and E.P. Scilingo, *PSYCHE: Personalised monitoring systems for care in mental health*, Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE, 2010, pp. 3602–3605. 1.2.1.2

[PCM+06] T. Pavani, G. Costa, M. Mazzotti, A. Conti, and D. Dardari, *Experimental Results on Indoor Localization Techniques through Wireless Sensors Network*, Vehicular Technology Conference, 2006. VTC 2006-Spring. IEEE 63rd, vol. 2, 2006, pp. 663–667. 2.2

[PFKP05] D.J. Patterson, D. Fox, H. Kautz, and M. Philipose, *Fine-grained activity recognition by aggregating abstract object usage*, Wearable Computers, 2005. Proceedings. Ninth IEEE International Symposium on, 2005, pp. 44–51. 4.2.3

[PFP+04] M. Philipose, K.P. Fishkin, M. Perkowitz, D.J. Patterson, D. Fox, H. Kautz, and D. Hahnel, *Inferring activities from interactions with objects*, Pervasive Computing, IEEE **3** (2004), no. 4, 50 – 57. 4.2.3, 4.5.1

[PH08] Brandon Paulson and Tracy Hammond, *Office activity recognition using hand posture cues*, Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction - Volume 2 (Swinton, UK, UK), BCS-HCI '08, British Computer Society, 2008, pp. 75–78. 1.2.1.5

[PL12] Gerald Pirkl and Paul Lukowicz, *Robust, low cost indoor positioning using magnetic resonant coupling*, Proceedings of the 2012 ACM Conference on Ubiquitous Computing (New York, NY, USA), UbiComp '12, ACM, 2012, pp. 431–440. 2.2

[POP98] C.P. Papageorgiou, M. Oren, and T. Poggio, *A general framework for object detection*, Computer Vision, 1998. Sixth International Conference on, 1998, pp. 555–562. 1.2.2

[PP00] Constantine Papageorgiou and Tomaso Poggio, *A Trainable System for Object Detection*, Int. J. Comput. Vision **38** (2000), no. 1, 15–33. 1.2.2

[PR12] Hamed Pirsiavash and Deva Ramanan, *Detecting Activities of Daily Living in First-person Camera Views*, Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE, 2012. 4.5.1

[PRA08] Shwetak N. Patel, Matthew S. Reynolds, and Gregory D. Abowd, *Detecting Human Movement by Differential Air Pressure Sensing in HVAC System Ductwork: An Exploration in Infrastructure Mediated Sensing.*, Pervasive (Jadwiga Indulska, Donald J. Patterson, Tom Rodden, and Max Ott, eds.), Lecture Notes in Computer Science, vol. 5013, Springer, 2008, pp. 1–18. 3.2

[Pri05] Nissanka Bodhi Priyantha, *The cricket indoor location system*, Tech. report, 2005. 2.2

[PRK+07] Shwetak N. Patel, Thomas Robertson, Julie A. Kientz, Matthew S. Reynolds, and Gregory D. Abowd, *At the flick of a switch: Detecting and classifying unique electrical events on the residential power line*, Proceedings of the 9th international conference on Ubiquitous computing (Berlin, Heidelberg), UbiComp '07, Springer-Verlag, 2007, pp. 271–288. 3.2.2, 3.3, 3.3.2, 3.5.3, 3.5.4

[PS12] N. Petersen and D. Stricker, *Learning task structure from video examples for workflow tracking and authoring*, Mixed and Augmented Reality (ISMAR), 2012 IEEE International Symposium on, 2012, pp. 237–246. 4.5.7

[PSKL08] G. Pirkl, K. Stockinger, K. Kunze, and P. Lukowicz, *Adapting magnetic resonant coupling based relative positioning technology for wearable activitiy recogniton*, Wearable Computers, 2008. ISWC 2008. 12th IEEE International Symposium on, 2008, pp. 47–54. 1.2.1.2, 1.2.1.3, 1.2.2, 1.2, 4.2.4

[PWF09] F.A. Pavel, Zhiyong Wang, and D.D. Feng, *Reliable object recognition using SIFT features*, Multimedia Signal Processing, 2009. MMSP '09. IEEE International Workshop on, 2009, pp. 1–6. 1.2.2, 4.2.2, 4.5.1.3

[RA06] M. S. Ryoo and J.K. Aggarwal, *Recognition of Composite Human Activities through Context-Free Grammar Based Representation*, Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, vol. 2, 2006, pp. 1709–1718. 4.10, 7.2

[RAG+01] Heather Richter, Gregory D. Abowd, Werner Geyer, Ludwin Fuchs, Shahrokh Daijavad, and Steven Poltrock, *Integrating Meeting Capture within a Collaborative Team Environment*, Ubicomp 2001: Ubiquitous Computing (Gregory D. Abowd, Barry Brumitt, and Steven Shafer, eds.), Lecture Notes in Computer Science, vol. 2201, Springer Berlin Heidelberg, 2001, pp. 123–138 (English). 1.2.1.5

[RDG+08] Jan S. Rellermeyer, Michael Duller, Ken Gilmer, Damianos Maragkos, Dimitrios Papageorgiou, and Gustavo Alonso, *The Software Fabric for the Internet of Things*, The Internet of Things: First International Conference, IOT 2008, Zurich, Switzerland, 2008, pp. 87–104. 6.5.3

[RFC+09] D. Roggen, K. Forster, A. Calatroni, T. Holleczek, Yu Fang, G. Troster, P. Lukowicz, G. Pirkl, D. Bannach, K. Kunze, A. Ferscha, C. Holzmann, A. Riener, R. Chavarriaga, and J. del R Millan, *OPPORTUNITY: Towards opportunistic activity and context recognition systems*, World of Wireless, Mobile and Multimedia Networks Workshops, 2009. WoWMoM 2009. IEEE International Symposium on a, 2009, pp. 1–6. 4.2.7

[RHM+02] Mandar A. Rahurkar, John H. L. Hansen, James Meyerhoff, George Saviolakis, and Michael Koenig, *Frequency band analysis for stress detection using a teager energy operator based feature*, INTERSPEECH'02, 2002, pp. –1–1. 5.2, 5.3, 5.8

[RK94] J.M. Rehg and T. Kanade, *DigitEyes: Vision-based hand tracking for human-computer interaction*, Motion of Non-Rigid and Articulated Objects, 1994., Proceedings of the 1994 IEEE Workshop on, 1994, pp. 16–22. 4.5.7

[RM00] Cliff Randell and H. Muller, *Context awareness by analysing accelerometer data*, Wearable Computers, The Fourth International Symposium on, 2000, pp. 175–176. 1.2.1.3, 2.3, 2.5.2, 4.2.1, 4.3, 4.5.3.2

[RSL12] A. Reiss, D. Stricker, and I. Lamprinos, *An Integrated Mobile System for Long-Term Aerobic Activity Monitoring and Support in Daily Life*, Trust, Security and Privacy in Computing and Communications (TrustCom), 2012 IEEE 11th International Conference on, 2012, pp. 2021–2028. 1.2.1.2

[RVC+07] R.P.D. Redondo, A.F. Vilas, M.R. Cabrer, J.J. Pazos Arias, and M.R. Lopez, *Enhancing Residential Gateways: OSGi Services Composition*, Consumer Electronics, 2007. ICCE 2007. Digest of Technical Papers. International Conference on, 2007, pp. 1–2. 6.2

[RVC+08] Rebeca P. Diaz Redondo, Ana Fernandez Vilas, Manuel Ramos Cabrer, Jose J. Pazos Arias, Jorge Garcia Duque, and Alberto Gil-Solla, *Enhancing Residential Gateways: A Semantic OSGi Platform.*, IEEE Intelligent Systems **23** (2008), no. 1, 32–40. 6.5.3

[RWHT11] Daniel Roggen, Martin Wirz, Dirk Helbing, and Gerhard Tröster, *Recognition of Crowd Behavior from Mobile Sensors with Pattern Analysis and Graph Clustering Methods*, Networks and Heterogeneous Media (2011). 1.2.1.6

[RYNDL11] LiKamWa Robert, Liu Yunxin, Lane Nicholas D., and Zhong Lin, *Can Your Smartphone Infer Your Mood?*, in PhoneSense workshop, 2011. 1.2.1.2, 1.2.2, 1.2

[SAW94] B. Schilit, N. Adams, and R. Want, *Context-Aware Computing Applications*, Proceedings of the 1994 First Workshop on Mobile Computing Systems and Applications (Washington, DC, USA), WMCSA '94, IEEE Computer Society, 1994, pp. 85–90. 1.1.1

[SBGP04] Adam Smith, Hari Balakrishnan, Michel Goraczko, and Nissanka Bodhi Priyantha, *Tracking Moving Devices with the Cricket Location System*, 2nd International Conference on Mobile Systems, Applications and Services (Mobisys 2004) (Boston, MA), June 2004. 2.2

[SCB11] R.R. Singh, S. Conjeti, and R. Banerjee, *An approach for real-time stress-trend detection using physiological signals in wearable computing systems for automotive drivers*, Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on, 2011, pp. 1477–1482. 5.2, 5.3

[SFR08] Jaakko Suutala, Kaori Fujinami, and Juha Röning, *Gaussian Process Person Identifier Based on Simple Floor Sensors*, Proceedings of the 3rd European Conference on Smart Sensing and Context (Berlin, Heidelberg), EuroSSC '08, Springer-Verlag, 2008, pp. 55–68. 2.2

[SHVLS08] M. Stikic, T. Huynh, K. Van Laerhoven, and B. Schiele, *ADL recognition based on the combination of RFID and accelerometer sensing*, Pervasive Computing Technologies for Healthcare, 2008. PervasiveHealth 2008. Second International Conference on, 2008, pp. 258–263. 4.2.3, 4.2.7

[SJN37] S. S. Stevens, Je, and E. B. Newman, *A scale for the measurement of the psychological magnitude of pitch*, J. Acoust Soc Amer **8** (1937), 185–190. 3.4.2.1

[SL07] D. Savio and T. Ludwig, *Smart Carpet: A Footstep Tracking Interface*, Advanced Information Networking and Applications Workshops, 2007, AINAW '07. 21st International Conference on, vol. 2, 2007, pp. 754–760. 4.7.3.1

[SL08] Axel Steinhage and Christl Lauterbach, *Monitoring Movement Behavior by Means of a Large Area Proximity Sensor Array in the Floor*, BMI'08, 2008, pp. 15–27. 2.2

[SLT04] Mathias Stäger, Paul Lukowicz, and Gerhard Tröster, *Implementation and Evaluation of a Low-Power Sound-Based User Activity Recognition System.*, ISWC, IEEE Computer Society, 2004, pp. 138–141. 3.2.1, 3.3, 3.3.1

[SLT07] Mathias Stäger, Paul Lukowicz, and Gerhard Tröster, *Power and accuracy trade-offs in sound-based context recognition systems*, Pervasive and Mobile Computing **3** (2007), no. 3, 300–327. 4.2.7

[SRE+05] J. Sivic, B.C. Russell, A.A. Efros, A. Zisserman, and W.T. Freeman, *Discovering objects and their location in images*, Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, vol. 1, 2005, pp. 370–377 Vol. 1. 1.2.2, 4.1.1.1, 4.2.2

[SRO+08] T. Stiefmeier, D. Roggen, G. Ogris, P. Lukowicz, and P. Lukowicz, *Wearable activity tracking in car manufacturing*, Pervasive Computing, IEEE **7** (2008), no. 2, 42–50. 4.3

[SSP98] Thad Starner, Bernt Schiele, and Alex Pentland, *Visual contextual awareness in wearable computing*, In International Symposium on Wearable Computing, 1998, pp. 50–57. 2.2, 4.2.2

[ST94] B.N. Schilit and M.M. Theimer, *Disseminating active map information to mobile hosts*, Network, IEEE **8** (1994), no. 5, 22–32. 1.1.2

[Sti08] Thomas Stiefmeier, *Real-Time Spotting of Human Activities in Industrial Environments*, Ph.D. thesis, ETH Zurich, 0 2008. 4.2.1, 4.3

[STS+13] Miguel Sousa, Axel Techmer, Axel Steinhage, Christl Lauterbach, and Paul Lukowicz, *Human Tracking and Identification using a Sensitive Floor and Wearable Accelerometers*, Pervasive Computing and Communications Workshops (PERCOM Workshops), 2013 IEEE International Conference on, 2013. 2.2

[SVLS08] M. Stikic, K. Van Laerhoven, and B. Schiele, *Exploring semi-supervised and active learning for activity recognition*, Wearable Computers, 2008. ISWC 2008. 12th IEEE International Symposium on, 2008, pp. 81–88. 1.2.2, 4.1.1.1

[TBB09] M. Tenorth, J. Bandouch, and M. Beetz, *The TUM Kitchen Data Set of everyday manipulation activities for motion tracking and action recognition*, Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on, 2009, pp. 1089–1096. 1.2.2

[TdOV08] Andre L. C. Tavares and Marco Tulio de Oliveira Valente, *A gentle introduction to OSGi.*, ACM SIGSOFT Software Engineering Notes **33** (2008), no. 5. 6.1, 6.5.1

[Ten00] David Tennenhouse, *Proactive computing*, Communications of the ACM **43** (2000), 43–50. 1.2.1.1

[TIL04] Emmanuel Munguia Tapia, Stephen S. Intille, and Kent Larson, *Activity Recognition in the Home using Simple and Ubiquitous Sensors*, In Pervasive, 2004, pp. 158–175. 4.2.5

[TJS10] Thiago Teixeira, Deokwoo Jung, and Andreas Savvides, *Tasking networked CCTV cameras and mobile phones to identify and localize multiple people*, Proceedings of the 12th ACM international conference on Ubiquitous computing (New York, NY, USA), Ubicomp '10, ACM, 2010, pp. 213–222. 2.2, 2.3, 2.8, 7.2

[TKSD11] Takumi Toyama, Thomas Kieninger, Faisal Shafait, and Andreas Dengel, *Museum Guide 2.0 - An Eye-Tracking based Personal Assistant for Museums and Exhibits*, Re-Thinking Technology in Museums 2011: Emerging Experiences. International Conference on Re-Thinking Technology in Museums, May 26-27, Limerick, Ireland (L. Ciolfi, K. Scott, and S. Barbieri, eds.), University of Limerick, 5 2011. 1.2.1.2

[TKSD12] T. Toyama, T. Kieninger, F. Shafait, and A. Dengel, *Gaze guided object recognition using a head-mounted eye tracker*, Proceedings of the Symposium on Eye Tracking Research and Applications, ACM, 2012, pp. 91–98. 4.2.2

[TMN+12] Ngo Thanh Trung, Y. Makihara, H. Nagahara, Y. Mukaigawa, and Y. Yagi, *Inertial-sensor-based walking action recognition using robust step detection and inter-class relationships*, Pattern Recognition (ICPR), 2012 21st International Conference on, 2012, pp. 3811–3814. 4.7.3.1

[TSGM10] B. Taati, J. Snoek, D. Giesbrecht, and A. Mihailidis, *Water Flow Detection in a Handwashing Task*, Computer and Robot Vision (CRV), 2010 Canadian Conference on, 2010, pp. 175–182. 3.2.1, 7.2

[VF02] D. Valtchev and I. Frankov, *Service gateway architecture for a smart home*, Communications Magazine, IEEE **40** (2002), no. 4, 126–132. 6.2, 6.5.3

[VLC00] K. Van Laerhoven and O. Cakmakci, *What shall we teach our pants*, Proceedings of the 4th IEEE International Symposium on Wearable Computers (ISWC), Citeseer, 2000, pp. 77–83. 1.2.1.3, 2.3, 4.2.1, 4.3, 4.3.1.1, 4.5.3.2, 4.7.3.1

[VSN+11] Trang Thuy Vu, A. Sokan, H. Nakajo, K. Fujinami, J. Suutala, P. Siirtola, T. Alasalmi, A. Pitkanen, and J. Roning, *Feature Selection and Activity Recognition to Detect Water Waste from Water Tap Usage*, Embedded and Real-Time Computing Systems and Applications (RTCSA), 2011 IEEE 17th International Conference on, vol. 2, 2011, pp. 138–141. 3.2.1, 7.2

[WDJ10] Hou Weifeng, Hou Di, and Chen Juan, *An OSGi Based RFID Complex Event Processing System*, Embedded and Ubiquitous Computing (EUC), 2010 IEEE/IFIP 8th International Conference on, 2010, pp. 162–169. 6.2

[Wei96] M. Weiser, *Open House*, In Review, the web magazine of the Interactive Telecommunications Program of New York University. Itp review 2.0., 1996. 1

[Wei99] Mark Weiser, *The computer for the 21st century*, SIGMOBILE Mob. Comput. Commun. Rev. **3** (1999), no. 3, 3–11. 1.1.3, 1.4.2

[WFMK+12] Martin Wirz, Tobias Franke, Eve Mitleton-Kelly, Daniel Roggen, Paul Lukowicz, and Gerhard Tröster, *CoenoSense: A framework for real-time detection and visualization of collective behaviors in human crowds by tracking mobile devices*, Proceedings of European Conference on Complex Systems, Springer, 2012. 1.2.1.6

[WFR+12] Martin Wirz, Tobias Franke, Daniel Roggen, Eve Mitleton-Kelly, Paul Lukowicz, and Gerhard Tröster, *Inferring and visualizing crowd conditions by collecting GPS location traces from pedestrians' mobile phones for real-time crowd monitoring during city-scale mass gatherings*, Collaborative Technology for Coordinating Crisis Management (CT2CM) track of WETICE-2012, 2012. 1.2.1.6, 5.3

[WHY09] Xiaoyu Wang, T.X. Han, and Shuicheng Yan, *An HOG-LBP human detector with partial occlusion handling*, Computer Vision, 2009 IEEE 12th International Conference on, 2009, pp. 32–39. 1.2.2

[WLG11] Jamie A. Ward, Paul Lukowicz, and Hans-Werner Gellersen, *Performance metrics for activity recognition.*, ACM TIST **2** (2011), no. 1, 6. 4.5.6.2, 4.7.8, 4.7.8.1, 4.7.8.2

[WLTS06a] J.A. Ward, P. Lukowicz, G. Troster, and T.E. Starner, *Activity Recognition of Assembly Tasks Using Body-Worn Microphones and Accelerometers*, Pattern Analysis and Machine Intelligence, IEEE Transactions on **28** (2006), no. 10, 1553–1567. 4.2.7

[WLTS06b] Jamie A. Ward, Paul Lukowicz, Gerhard Tröster, and Thad Starner, *Activity recognition of assembly tasks using body-worn microphones and accelerometers*, IEEE Transactions on Pattern Analysis and Machine Intelligence **28** (2006), 2006. 1.2.1.1

[WNS06] C. Wojek, K. Nickel, and R. Stiefelhagen, *Activity Recognition and Room-Level Tracking in an Office Environment*, Multisensor Fusion and Integration for Intelligent Systems, 2006 IEEE International Conference on, 2006, pp. 25–30. 4.2.7

[WOC+07] Jianxin Wu, Adebola Osuntogun, Tanzeem Choudhury, Matthai Philipose, and James M. Rehg, *A Scalable Approach to Activity Recognition based on Object Use*, Proc. of ICCV07, 2007, pp. 1–8. 4.2.3

[WPP+07] Shiaokai Wang, William Pentney, Ana-Maria Popescu, Tanzeem Choudhury, and Matthai Philipose, *Common sense based joint training of human activity recognizers*, Proceedings of the 20th international joint conference on Artifical intelligence (San Francisco, CA, USA), IJCAI'07, Morgan Kaufmann Publishers Inc., 2007, pp. 2237–2242. 4.2.7

[WPVS05] Gertjan Wijnalda, Steffen Pauws, Fabio Vignoli, and Heiner Stuckenschmidt, *A Personalized Music System for Motivation in Sport Performance*, IEEE Pervasive Computing **4** (2005), no. 3, 26–32. 1.2.1.3

[WSP+11] Martin Wirz, Christina Strohrmann, Roman Patscheider, Fabian Hilti, Bernhard Gahr, Frederik Hess, Daniel Roggen, and Gerhard Tröster, *Real-time detection and recommendation of thermal spots by sensing collective behaviors in paragliding*, Proceedings of 1st international symposium on From digital footprints to social and community intelligence (New York, NY, USA), SCI '11, ACM, 2011, pp. 7–12. 1.2.1.3

[WWS07] Jie Wu, Dong Wang, and Huanye Sheng, *Design an OSGi Extension Service for Mobile RFID Applications*, e-Business Engineering, 2007. ICEBE 2007. IEEE International Conference on, 2007, pp. 323–326. 6.2

[YCKV07] Tao Yang, F. Chen, D. Kimber, and J. Vaughan, *Robust People Detection and Tracking in a Multi-Camera Indoor Visual Surveillance System*, Multimedia and Expo, 2007 IEEE International Conference on, 2007, pp. 675–678. 4.7.3.1

[YOI92] J. Yamato, Jun Ohya, and K. Ishii, *Recognizing human action in time-sequential images using hidden Markov model*, Computer Vision and Pattern Recognition, 1992. Proceedings CVPR '92., 1992 IEEE Computer Society Conference on, 1992, pp. 379–385. 4.2.2

[YYA+10] Zhiwen Yu, Zhiyong Yu, H. Aoyama, M. Ozeki, and Y. Nakamura, *Capture, recognition, and visualization of human semantic interactions in meetings*, Pervasive Computing and Communications (PerCom), 2010 IEEE International Conference on, 2010, pp. 107–115. 1.2.1.5

[Zö9] Detlef Zühlke, *SmartFactory - A Vision becomes Reality*, Keynote Papers of the 13th IFAC Symposium on Information Control Problems in Manufacturing (INCOM 09). IFAC Symposium on Information Control Problems in Manufacturing (INCOM-09), 13th, June 3-5, Moscow, Russian Federation, ICS / RAS, 6 2009. 1.2.2

[ZBMM06] Hao Zhang, A.C. Berg, M. Maire, and J. Malik, *SVM-KNN: Discriminative Nearest Neighbor Classification for Visual Category Recognition*, Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, vol. 2, 2006, pp. 2126–2136. 1.2.2, 4.1.1.1

[ZBS09] Andreas Zinnen, Ulf Blanke, and Bernt Schiele, *An Analysis of Sensor-Oriented vs. Model-Based Activity Recognition*, Proceedings of the 13th IEEE International Symposium on Wearable Computers (ISWC), 2009. 4.1.1.3

[ZDC11] H. Zhang, R. Dantu, and J. W. Cangussu, *Socioscope: Human Relationship and Behavior Analysis in Social Networks*, Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on **41** (2011), no. 6, 1122–1143. 1.2.1.5, 5.2, 5.3

[ZMTA09] J.Z. Zhang, N. Mbitiru, P.C. Tay, and R.D. Adams, *Analysis of Stress in speech using adaptive Empirical Mode Decomposition*, Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on, nov. 2009, pp. 361–365. 5.2, 5.3, 5.8

[ZS08] A. Zinnen and B. Schiele, *A new approach to enable gesture recognition in continuous data streams*, Wearable Computers, 2008. ISWC 2008. 12th IEEE International Symposium on, 2008, pp. 33–40. 4.5.3

[ZS11] Chun Zhu and Weihua Sheng, *Motion- and location-based online human daily activity recognition*, Pervasive Mob. Comput. **7** (2011), no. 2, 256–269. 2.2, 2.3

[ZS13] M. Zhang and A. Sawchuk, *Human Daily Activity Recognition with Sparse Representation Using Wearable Sensors*, vol. PP, 2013, pp. 1–1. 4.2.1, 4.3

[ZUA02] Jörg Zieren, Nils Unger, and Suat Akyol, *Hands Tracking from Frontal View for Vision-Based Gesture Recognition*, Proceedings of the 24th DAGM Symposium on Pattern Recognition (London, UK, UK), Springer-Verlag, 2002, pp. 531–539. 4.5.7

[ZVLS07] Andreas Zinnen, Kristof Van Laerhoven, and Bernt Schiele, *Toward Recognition of Short and Non-repetitive Activities from Wearable Sensors*, AmI, Lecture Notes in Computer Science, vol. 4794, Springer, Springer, 2007, pp. 142–158. 4.3.1.1

[ZWS09] Andreas Zinnen, Christian Wojek, and Bernt Schiele, *Multi Activity Recognition based on Bodymodel-Derived Primitives*, 4th International Symposium on Location and Context Awareness (LoCA) (Tokyo, Japan), 2009. 1.2.1.1, 1.2.2, 4.3, 4.5.3

[ZZS07] Wei Zhang, G. Zelinsky, and D. Samaras, *Real-time Accurate Object Detection using Multiple Resolutions*, Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, 2007, pp. 1–8. 4.2.2, 4.5.1.3, 4.5.5, 4.7.1.1, 4.7.1.2, 4.7.1.2

Curriculum Vitae

Gerald Bauer
Email: thesis@gerald-bauer.de

Education

2012 - 2013	PhD student at the German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany.
2007 - 2012	PhD student at the Embedded Systems Lab, University of Passau, Germany.
2001 - 2007	Studies of computer science, graduation with degree *Dipl.-Inf.*
	Major subjects: Intelligent technical systems, algorithms and databases.
	Minor subjects: Media and Design (Specialization: Cyberlaw).
	Diploma thesis: Klassifikation von Radar- und Laserscanner-Rohdaten im Bereich der Fahrumfelderfassung, supervised by Dr. Thomas Tatschke and Prof. Dr. Klaus Donner (FORWISS, Passau, Germany).

Experience

2012 - 2013	Founding member of SIS (Social Information Solution) Software GmbH. Position: Head of Emerging Technologies Lab
2012 - 2013	Research and management position at the Embedded Intelligence Group, DFKI, Kaiserslautern, Germany.

2010 - 2012	FP7 EU project INTERSTRESS - Interreality in the Management and Treatment of Stress-Related Disorders: Leader of work package "Behavioral Parameters".
2008 - 2010	Several research projects abroad in cooperation with the Tecnodiscap Group, University of Zaragoza, Spain.
2007 - 2011	FP6 EU project MonAMI - Mainstreaming on Ambient Intelligence: Leader of work package "Monitoring".
2007 - 2013	Researcher at the Embedded System Lab (University of Passau; 2007-2012) and at the German Research Center for Artificial Intelligence (DFKI, Kaiserslautern; 2012-2013): Pervasive and wearable computing.

Publications

The following table shows a selection of my publications that built the basis of this thesis.

Chapter	Title
2	Gerald Bauer and Paul Lukowicz. *Developing a Sub Room Level Indoor Location System for Wide Scale Deployment in Assisted Living Systems*, ICCHP (K. Miesenberger, J. Klaus, W. L. Zagler & A. I. Karshmer, eds.), Lecture Notes in Computer Science, vol. 5105, Springer, 2008, pp. 1057-1064.
3	Alejandro Ibarz, Gerald Bauer, Roberto Casas, Alvaro Marco, and Paul Lukowicz. *Design and Evaluation of a Sound Based Water Flow Measurement System*, EuroSSC (Daniel Roggen, Clemens Lombriser, Gerhard Tröester, Gerd Kurtuem, and Paul J.M. Havinga, eds.), Lecture Notes in Computer Science, vol. 5279, Springer, 2008, pp. 41-54. – ***Best Commercial Potential Award*** Gerald Bauer, Karl Stockinger and Paul Lukowicz. *Recognizing the Use-Mode of Kitchen Appliances from their Current Consumption*, EuroSSC (Payam M. Barnaghi, Klaus Moessner, Mirko Presser, and Stefan Meissner, eds.), Lecture Notes in Computer Science, vol. 5741, Springer, 2009, pp. 163-176. – ***Best Paper Award***
4	Gerald Bauer, Ulf Blanke, Bernt Schiele, and Paul Lukowicz. *User-Independent, Multi-Modal Spotting of Subtle Arm Actions with Minimal Training Data*, 10th IEEE Workshop on Context Modeling and Reasoning 2013, San Diego.
5	Gerald Bauer and Paul Lukowicz. *Can Smartphones Detect Stress-Related Changes in the Behavior of Individuals?*, in 'PerCom Workshops', IEEE, 2012, pp. 423-426.
6	Alvaro Marco, Roberto Casas, Gerald Bauer, Ruben Blasco Marin, Angel Asensio, Bruno Jean-Bart and Miriam Ibanez. *Common OSGi Interface for Ambient Assisted Living Scenarios*, in Björn Gottfried and Hamid K. Aghajan, ed., 'BMI Book', Volume 3, Ambient Intelligence and Smart Environments, IOS Press, 2009, pp. 336-357. Gunnar Fagerberg, Antonio Kung, Reiner Wichert, Mohammad-Reza Tazari, Bruno Jean-Bart, Gerald Bauer, Gottfried Zimmermann, Francesco Furfari, Francesco Potorti, Stefano Chessa, Michael Hellenschmidt, Joe Gorman, Jan Alexandersson, Jürgen Bund, Eduardo Carrasco, Gorka Epelde, Martin Klima, Elena Urdaneta, Gregg C. Vanderheiden and Ingo Zinnikus. *Platforms for AAL Applications*, in Paul Lukowicz; Kai S. Kunze and Gerd Kortuem, ed., 'EuroSSC', Springer, 2010, pp. 177-201.